Méthodes asymptotico-numériques pour des problèmes issus de la physique des plasmas et de la modélisation des interactions sociales

Laurent Navoret

THÈSE

En vue de l'obtention du

Doctorat de l'Université de Toulouse

Délivré par l'Université Toulouse 3 - Paul Sabatier
Discipline : Mathématiques

Présentée et soutenue par **Laurent Navoret**
Le 30 juin 2010

Titre :

Méthodes asymptotico-numériques pour des problèmes issus de la physique des plasmas et de la modélisation des interactions sociales

---

JURY :

Naoufel Ben Abdallah, Université Toulouse 3 (Examinateur)
Didier Bresch, Université de Savoie (Rapporteur)
Pierre Degond, Université Toulouse 3 (Directeur de thèse)
Bertrand Maury, Université Paris Sud (Rapporteur)
Frédéric Pascal, ENS de Cachan (Examinateur)
Olivier Pironneau, Université Paris 6 (Examinateur)
David Sanchez, INSA de Toulouse (Directeur de thèse)
Guy Theraulaz, Université Toulouse 3 (Examinateur)

Ecole Doctorale
Mathématiques Informatique Télécommunications de Toulouse

Laboratoire : **Institut de Mathématiques de Toulouse**
Université Toulouse 3 - Paul Sabatier 31062 TOULOUSE cedex 9, France

Remerciements

Je tiens en premier lieu à remercier chaleureusement mes deux directeurs de thèse, Pierre Degond et David Sanchez, qui m'ont proposé un sujet d'étude passionnant. Durant ces trois années de thèse, j'ai grandement apprécié l'enthousiasme de Pierre, ses conseils et sa disponibilité. Je suis aussi très reconnaissant envers David pour son aide. Merci encore à tous les deux.

Je tiens à exprimer toute ma gratitude à Didier Bresch et Bertrand Maury qui ont accepté de rapporter ce manuscrit. Guy Theraulaz a été un interlocuteur privilégié pendant cette thèse et c'est un plaisir pour moi qu'il fasse partie de mon jury. J'ai beaucoup apprécié le travail de Frédéric Pascal à l'Ecole Normale Supérieure de Cachan et je suis donc très heureux qu'il ait accepté de présider mon jury de thèse. Je remercie également Olivier Pironneau qui me fait l'honneur de faire partie de mon jury. Enfin, j'ai une pensée émue pour Naoufel Ben Abdallah qui vient de nous quitter. Merci pour votre gentillesse.

Je remercie Frédéric Lagoutière pour son énergie durant son cours d'EDP, printemps 2005, et pour sa passion communicative.

Durant ma thèse, j'ai eu le plaisir de travailler avec Fabrice Deluzet que je remercie pour sa précieuse aide, son efficacité et ses encouragements lorsque j'éprouvais des difficultés avec mon code Fortran. Sébastien Motsch, Marie-Hélène Vignal, Adrien Bonneu, An-Bang Sun et enfin Jiale Hua ont fortement contribué, par les multiples discussions que j'ai pu avoir avec eux, à rendre ces années de thèse passionnantes et je les en remercie.

Je remercie chaleureusement Richard Bon, Jacques Gautrais, Guy Theraulaz, Marie-Hélène Pillot et Mehdi Moussaïd, pour m'avoir fait découvrir un autre univers, celui de la recherche en biologie. Je remercie également Hugues Chaté et plus généralement tous les membres des projets Panurge et Pedigree pour nos débats captivants.

Je tiens à remercier l'ensemble des personnels de l'Institut de Mathématiques de Toulouse, chercheurs et non chercheurs, pour leur accueil.

Je remercie tous les collègues avec qui j'ai eu la joie de partager le bureau : je pense en tout premier lieu à Laetitia, la plus fidèle des co-bureaux, puis à Domi, Sever, Min et enfin Stefan. Que de beaux moments partagés à apprendre le roumain, la planche à voile, la spintronique et un peu de chinois !

Je remercie aussi Benjamin et Salvador pour les diners de chef Simon qui ont agrémenté mes soirées. Merci aussi à Amic et au cours de jongle, à la Complex Systems team (Robin, Manu, Chaker, Dario, Mathieu, Giacomo, Anaïs). Je dois aussi remercier la cafetière du bureau 202 et le préposé à l'achat du café qui se reconnaîtra. Merci aussi

au carom et à Léo, à Marion (les 2), Tiphaine, Mélanie, Jorge,... mais aussi merci aux “grands sages" : Laetitia, Michael, Afeintu, Jean-Luc, Raymond, Samy, Marc,...

Je salue aussi tous les membres de l'équipe EMCC. Les côtoyer a été réellement enrichissant.

Je remercie enfin mes parents et ma soeur pour leur soutien indéfectible. J'ai aussi une pensée particulière pour le reste de ma famille et pour tous les amis, toujours aussi indispensables.

Et pour finir, mes plus grands remerciements vont à Perrine qui m'a accompagné et soutenu dans cette aventure.

Résumé

Dans cette thèse, nous nous intéressons à différents problèmes asymptotiques issus de la physique des plasmas et de la modélisation des mouvements collectifs dans les populations animales. Nous nous attachons à développer des méthodes analytiques et numériques pour capturer les dynamiques asymptotiques. Le premier problème est la limite quasi-neutre dans les plasmas et les deux autres sont la limite hydrodynamique avec une contrainte géométrique et la limite de congestion dans des modèles biologiques de déplacement.

Dans une première partie, nous présentons une méthode numérique Particle-In-Cell (PIC) pour le système Vlasov-Poisson préservant l'asymptotique quasi-neutre. Contrairement aux méthodes classiques, elle n'est pas sujette aux sévères contraintes de stabilité induites par des faibles longueurs de Debye et des fréquences plasma élevées. C'est un schéma asymptotiquement stable et asymptotiquement consistant. Plusieurs cas tests unidimensionnels permettent de le valider et de le comparer avec d'autres schémas.

Dans la deuxième partie, nous nous intéressons à la description macroscopique du modèle de Vicsek, qui décrit un système de particules avec vitesse de norme constante et des interactions d'alignement. Le modèle macroscopique de Vicsek est un système hyperbolique non-conservatif avec une contrainte géométrique. Tout d'abord, nous présentons la limite macroscopique d'un modèle de Vicsek à deux populations, une population à l'arrêt et une population en mouvement, avec des échanges modélisant les passages de l'état immobile à l'état de déplacement. Pour valider le modèle de Vicsek macroscopique et capturer les solutions correspondant à la dynamique particulaire sous-jacente, plusieurs schémas numériques sont proposés et l'un d'entre eux, basé sur une relaxation de la contrainte géométrique, apporte de bons résultats.

La troisième partie est dédiée à l'étude des effets de congestion dans les modèles de déplacement. A partir d'un modèle particulaire d'attraction à longue portée et répulsion à courte portée, nous dérivons un modèle macroscopique hyperbolique non-conservatif avec une contrainte de densité maximale. En accentuant la raideur de cette contrainte, la transition entre les régions de densités maximales et celles de densités plus faibles se traduit asymptotiquement par une transition de type compressible-incompressible. L'analyse du problème de Riemann unidimensionnel permet d'établir en partie la dynamique de cette transition. Finalement, deux schémas numériques préservant l'asymptotique de congestion pour le système d'Euler avec contrainte de densité maximale sont proposés et comparés.

Table des matières

Introduction

Dans cette thèse, nous nous intéressons à différents problèmes asymptotiques issus de la physique des plasmas et de la modélisation des mouvements collectifs dans les populations animales. Cette introduction les présente dans leurs contextes respectifs. Nous décrivons brièvement les méthodes analytiques et numériques utilisées pour capturer les dynamiques asymptotiques puis nous présentons les résultats obtenus dans cette thèse.

1 Problèmes asymptotiques

1.1 Cadre - Définition

Les problèmes asymptotiques étudiés dans cette thèse sont liés à la modélisation de phénomènes physiques ou biologiques. Ils se présentent lorsque l'échelle temporelle ou spatiale d'une dynamique est bien plus petite que l'échelle de description voulue par l'observateur (en l'occurence nous). Nous sommes donc intéressés par une description simplifiée ou à grande échelle du phénomène. Mais avant de donner des exemples concrets, introduisons tout de suite le formalisme suivant :

Problème asymptotique. Nous nous intéressons à des problèmes dépendants d'un paramètre ε et à leurs limites lorsque ce paramètre tend vers 0. Nous noterons P^ε un tel problème et P^0 sa limite :

$$P^\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} P^0.$$

Les problèmes asymptotiques considérés dans cette thèse s'inscrivent dans l'étude des équations aux dérivées partielles. Présentons deux exemples typiques de problèmes asymptotiques.

1. Un gaz, au sein duquel interagissent un grand nombre de particules, est décrit par la fonction de distribution des particules $f(x, v, t)$ dans l'espace des phases, c'est-à-dire l'espace des positions $x \in \mathbb{R}^3$ et des vitesses $v \in \mathbb{R}^3$ et où $t \in \mathbb{R}^+$ désigne la variable de temps. Celle-ci vérifie l'équation de Boltzmann suivante :

$$(\partial_t + v \cdot \nabla_x) f = \frac{1}{\varepsilon} Q(f), \qquad (P_B^\varepsilon)$$

où l'opérateur $(\partial_t + v \cdot \nabla_x)$ modélise le transport des particules, l'opérateur $Q(f)$ les collisions entre particules. Le paramètre ε est ici le libre parcourt moyen ℓ adimensionné par l'échelle des longueurs L :

$$\varepsilon = \frac{\ell}{L} := Kn.$$

Il est appelé nombre de Knudsen et est noté Kn. L'opérateur $Q(f)$ agit seulement sur les distributions en vitesse et se comporte comme une **relaxation** vers les équilibres locaux (c'est-à-dire les distributions f telles que $Q(f) = 0$). Lorsque ε devient très petit (ℓ devient très petit par rapport à L), la limite de (P_B^ε) est le système d'Euler compressible qui donne la dynamique des paramètres des équilibres locaux, à savoir la densité, la quantité de mouvement et la température. Nous renvoyons le lecteur à la revue [40].

2. Le système d'Euler compressible isentropique est constitué des équations exprimant les conservations de la densité de masse $\rho(x,t) \geq 0$ et de la quantité de mouvement $q(x,t) = \rho u(x,t) \in \mathbb{R}^d$, avec $x \in \mathbb{R}^d$ et $t \in \mathbb{R}^+$:

$$\begin{aligned} &\partial_t \rho + \nabla_x \cdot q = 0, \\ &\partial_t q + \nabla_x \cdot \left(\frac{q \otimes q}{\rho} \right) + \frac{1}{\varepsilon^2} \nabla_x p(\rho) = 0, \end{aligned} \qquad (P_{BM}^\varepsilon)$$

où p désigne la pression et est définie par la loi d'état isentropique $p(\rho) = \rho^\gamma$, $\gamma > 1$. Pour un champ scalaire $a(x)$, $\nabla_x a$ désigne le gradient de a et pour un champ de vecteur $A = (A_i(x))_{i=1,..,d}$ ou un tenseur $B = (B_{i,j}(x))_{i,j=1,..,d}$, $\nabla_x \cdot A$ et $\nabla_x \cdot B$ désignent leurs divergences :

$$(\nabla_x a)_i = \partial_{x_i} a, \quad \nabla_x \cdot A = \sum_{i=1}^d \partial_{x_i} A_i, \quad (\nabla_x \cdot B)_i = \sum_{j=1}^d \partial_{x_j} B_{i,j}.$$

Enfin, le produit tensoriel $u \otimes v$ de deux vecteurs $u = (u_i)_{i=1,..,d}$ et $v = (v_i)_{i=1,..,d}$ est défini par : $(u \otimes v)_{i,j} = u_i v_j$.
Le paramètre ε est le nombre de Mach (il est noté aussi Ma). C'est le ratio de l'échelle des vitesses du fluide u_0 sur la vitesse de propagation des ondes acoustiques c_s (ou vitesse du son) :

$$\varepsilon = \frac{u_0}{c_s} := Ma.$$

Lorsque ε devient très petit, la vitesse des ondes acoustiques devient infinie et la densité s'uniformise. De plus, la vitesse u satisfait à la limite $\varepsilon \to 0$ la contrainte d'incompressibilité :

$$\nabla_x \cdot u = 0.$$

La limite du système (P_{BM}^ε) est le système d'Euler incompressible [74].

Echelle de description. Au vu des deux exemples précédents, le paramètre ε est lié à un changement d'échelle. Il peut être un paramètre physique adimensionné, tel que le nombre de Knudsen, le nombre de Mach, la longueur de Debye dans les plasmas (voir paragraphe 1.2), le nombre de Reynolds (caractérisant la viscosité d'un écoulement), ... Il caractérise alors une dynamique à une échelle donnée et la limite $\varepsilon \to 0$ traduit la raideur du problème mathématique associé vis-à-vis de l'échelle de l'observateur.

Le paramètre ε peut aussi être introduit comme un changement d'échelle en faisant un **changement des variables** d'espace et de temps. Nous utiliserons par exemple le changement d'échelle hydrodynamique :

$$t' = \varepsilon t, \quad x' = \varepsilon x, \quad \varepsilon \ll 1.$$

Validité de P^0. Rattacher ε à un paramètre physique permet de souligner le fait que le régime du problème asymptotique P^0 n'est pas toujours valide : si ε peut être très faible à certains endroits ($\varepsilon \ll 1$), il peut être aussi de l'ordre de l'unité à d'autres ($\varepsilon = O(1)$). Le paramètre ε peut donc varier en espace et en temps, $\varepsilon = \varepsilon(x,t)$, du fait de l'évolution des propriétés physiques du système considéré.

Pourquoi s'intéresse-t-on à P^0 ? Comme nous l'avons annoncé, P^0 procure une "**simplification**" du problème. Nous espérons en effet capturer seulement la dynamique macroscopique et les dynamiques à petites échelles ne seront perçues que par leurs effets moyens. De ce point de vue, P^0 se trouve être le modèle le plus approprié pour l'observateur. Il a l'avantage de permettre des simulations numériques moins coûteuses et d'être un outil pour classifier les dynamiques. Bien sûr, la modélisation d'un processus à grande échelle entraîne nécessairement une perte de précision.

Cette notion de "simplification" est néanmoins relative car les natures mathématiques de P^ε et P^0 peuvent être différentes : ceci implique notamment des défauts de convergence des solutions (problème de couche limite) ou une sous-détermination du problème (limite visqueuse pour les problèmes hyperboliques). La limite est dite singulière et nous dirons qu'il y a dégénérescence.

Les difficultés. Tout d'abord, la limite $P^\varepsilon \to P^0$ et donc la validité de P^0 dépend souvent d'hypothèses : la limite asymptotique est-elle bien pertinente ? Les réponses analytiques faisant souvent défaut, des simulations numériques peuvent donner quelques indices.

Une deuxième difficulté peut survenir lorsque la connaissance du problème asymptotique P^0 n'est que partielle. Les méthodes numériques que nous introduirons au paragraphe 2 sont alors un outil pour décrire les solutions du problème asymptotique.

Dans la suite, nous présentons les trois problèmes asymptotiques auxquels nous nous sommes intéressés : la limite quasi-neutre en physique des plasmas, la limite hydrodynamique avec contrainte géométrique et la limite de congestion pour des modèles de mouvements collectifs dans des populations animales.

1.2 Exemple 1 : Limite quasi-neutre

1.2.1 Contexte : la physique des plasmas

Un plasma est un gaz ionisé, c'est-à-dire contenant des particules chargées, électrons et ions. Le plasma est l'état de la matière le plus répandu dans l'Univers : l'intérieur des étoiles, le vent solaire sont des plasmas et plus près de nous l'ionosphère et les tubes à néons en contiennent. La production d'énergie électrique par fusion nucléaire nécessite aussi l'utilisation de plasmas.

Ions et électrons sont soumis aux forces coulombiennes dues au champ électromagnétique créé par leur propre distribution : les forces sont donc non locales et donnent lieu à des comportements collectifs. De plus, du fait des interactions complexes entre les ondes électro-magnétiques et les particules, la physique des plasmas est particulièrement riche : les relations de dispersions des ondes plasmas dépendent des directions relatives

des champs électriques et magnétiques, des instabilités doubles faisceaux se produisent dues à des effets Doppler, des amortissements de type Landau ont lieu contredisant à première vue le caractère réversible d'une dynamique non-collisionnelle... Nous renvoyons à [30] pour plus détails.

Dans la suite, nous considèrerons des plasmas totalement ionisés (sans particules neutres), c'est-à-dire des plasmas chauds, et non collisionnels.

1.2.2 Longueur de Debye et quasi-neutralité

Lorsqu'une charge positive est introduite au sein du plasma parfaitement neutre avec un certain potentiel, celle-ci est immédiatement entourée par un nuage de charges négatives de telle sorte que la charge effective locale s'annule et que le potentiel créé par la chare positive soit négligeable en dehors du nuage. Cet effet est appelé écrantage de Debye. La taille de ce nuage mesure l'échelle de longueur des déséquilibres de charges au sein du plasma. Elle est donnée par la **longueur de Debye** :

$$\lambda_D = \sqrt{\frac{k_B T_0 \varepsilon_0}{e^2 n_0}},$$

où k_B désigne la constante de Boltzmann, T_0 est l'échelle de température des électrons, ε_0 la permittivité diélectrique et n_0 l'échelle de densité de particules. Elle correspond aux lieux où énergies thermiques $k_B T_0$ et potentielles $e^2(n_0\lambda_D^3)/(\varepsilon_0\lambda_D)$ sont identiques. Notons que le caractère non-collisionnel du plasma nécessite de plus que le nombre de particules au sein de la sphère de Debye soit grand pour que les effets collectifs prédominent.

La quasi-neutralité a lieu lorsque $\lambda_D \ll L$, où L est la longueur caractéristique du problème. Cette quasi-neutralité peut néanmoins être brisée localement, notamment aux interfaces plasma-vide : on parle alors de gaine.

L'échelle temporelle caractéristique est celle du retour à la quasi-neutralité et par conséquent à la propagation des défauts de neutralité. Les électrons agissent comme des ressorts, dont la fréquence d'oscillation est donnée par la **fréquence plasma** ω_p :

$$\omega_p = \sqrt{\frac{n_0 e^2}{\varepsilon_0 m_e}}$$

où m_e désigne la masse des électrons. Combinées avec le mouvement thermique des électrons, ces oscillations induisent la formation d'ondes électroniques et ioniques.

1.2.3 Le système Vlasov-Poisson et la quasi-neutralité

Pour simplifier la présentation, nous considèrerons que les ions sont immobiles et sont distribués de manière homogène, avec une densité égale à 1. La dynamique des électrons est décrite par la fonction de distribution $f(x, v, t)$ où $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$, avec $d = 1, 2$ ou 3, sont respectivement les variables de position et de vitesse et $t \in \mathbb{R}^+$ désigne le temps. Cette fonction de distribution vérifie, après adimensionnement, le système Vlasov-Poisson suivant :

$$\partial_t f + v \cdot \nabla_x f + \nabla_x \phi \cdot \nabla_v f = 0, \tag{1}$$
$$-\varepsilon^2 \Delta_x \phi = 1 - n, \tag{2}$$

où f satisfait l'équation de Vlasov (1), le potentiel électrique ϕ satisfait l'équation de Poisson (2) et $n(x,t)$ désigne la densité d'électrons

$$n(x,t) = \int_{\mathbb{R}^d} f(x,v,t)dv.$$

$u \cdot v$ désigne le produit scalaire de deux vecteurs u et v et $\Delta_x \phi$ désigne le laplacien de ϕ ($\Delta_x \phi = \sum_{j=1,..,d} \partial_{x_j^2} \phi$). Le paramètre $\varepsilon = \lambda_D / L$ est la longueur de Debye adimensionnée. L'existence et l'unicité des solutions du système Vlasov-Poisson a fait l'objet de nombreuses études. Nous ne détaillons pas ici toutes les contributions. Mentionnons toutefois les travaux [3] pour ce qui concerne les solutions faibles, et [86, 101] pour les solutions fortes.

Lorsque ε tend vers 0, l'équation de Poisson dégénère au sens où l'équation de Poisson à l'ordre $O(1)$ n'est pas une équation elliptique mais une égalité algébrique :

$$n = 1.$$

Le potentiel est lui déterminé par les variations d'ordre $O(\varepsilon^2)$ de la densité.

Reformulation de l'équation de Poisson. Pour obtenir une équation qui ne dégénère pas, nous faisons appel aux équations sur les deux premiers moments de f, à savoir la densité n et la quantité de mouvement nu définies par :

$$n(x,t) = \int_{\mathbb{R}^d} f(x,v,t)dv, \quad nu(x,t) = \int_{\mathbb{R}^d} f(x,v,t)vdv.$$

Par intégration de l'équation de Vlasov, nous obtenons les équations d'Euler :

$$\partial_t n + \nabla_x \cdot (nu) = 0, \tag{3}$$
$$\partial_t (nu) + \nabla_x \cdot S = n \nabla_x \phi, \tag{4}$$

où $S = \int_{\mathbb{R}^d} f(v \otimes v)dv$ est le flux de nu convectif. En soustrayant la dérivée temporelle de (3) à la dérivée spatiale de (4) et utilisant l'équation de Poisson, nous obtenons l'équation de Poisson reformulée :

$$-\nabla_x \cdot \left(\left(n + \varepsilon^2 \partial_{tt}^2 \right) \nabla_x \phi \right) = -\nabla_x^2 : S,$$

où ∇_x^2 désigne le tenseur des dérivées d'ordre deux ($(\nabla_x^2)_{i,j} = \partial_{x_i x_j}$) et si $A = (A_{i,j})_{i,j=1,..,d}$ et $B = (B_{i,j})_{i,j=1,..,d}$ sont deux tenseurs, $A : B$ désigne leur produit contracté ($A : B = \sum_{i,j=1,..,d} A_{i,j} B_{i,j}$). Cette équation montre notamment que le champ électrique $\nabla_x \phi$ est fortement oscillant. C'est ce qui constitue la difficulté des études théoriques de ce problème asymptotique : les notions de solutions à valeurs mesures et de mesures de défauts ont été utilisées pour surmonter ce problème dans [58]. La convergence vers le système Euler incompressible pour les plasmas froids a été montrée dans [21, 88] à partir de la notion de solution dissipative du système d'Euler.

Le système Vlasov-Poisson est donc reformulé ainsi :

$$\begin{aligned} &\partial_t f + v \cdot \nabla_x f + \nabla_x \phi \cdot \nabla_v f = 0, \\ &-\nabla_x \cdot \left(\left(n + \varepsilon^2 \partial_{tt}^2 \right) \nabla_x \phi \right) = -\nabla_x^2 : S, \end{aligned} \tag{P_{VP}^{ε}}$$

Cette reformulation est à la base de la méthode numérique, que nous introduisons au paragraphe 2.2. Elle a déjà été utilisée dans le cadre du système Euler-Poisson [37].

Autres asymptotiques en physique des plasmas. Avant de clotûrer ce paragraphe, signalons que d'autres problèmes asymptotiques peuvent être considérés : la limite $m_e \to 0$ (les électrons sont dits Boltzmannien), la limite gyro-cinétique lorsque les particules sont soumises à un fort champ magnétique [21], la limite bas-Mach lorsque la vitesse du plasma est petite devant la vitesse du son [41].

1.3 Exemple 2 : Limite hydrodynamique avec contrainte géométrique

1.3.1 Contexte : mouvement collectif dans les populations animales

Nous nous intéressons à présent à la description à grande échelle des mouvements collectifs observés dans les populations animales, telles que les colonies de bactéries, les essaims d'insectes [18] (fourmis, criquets, etc.), les bancs de poissons [98], les nuées d'oiseaux [6] ou encore les troupeaux de mammifères [34] (moutons, gnous) et les foules de piétons [65]. Nous renvoyons aux écrits [25, 110] pour une présentation plus générale des comportements collectifs et de l'auto-organisation dans le monde animal. Alors que le nombre d'individus dans un troupeaux de moutons est l'ordre du millier au plus, un troupeau de gnous ou une colonie de fourmis peut compter de l'ordre de 100 000 individus [34, 47]. Il peut donc s'avérer nécessaire de développer des modèles permettant de rendre compte de manière synthétique de la multitude des interactions. Pour cela, une méthode consiste à se donner des lois d'interactions entre les particules, puis de dériver un modèle macroscopique par un changement d'échelle.

Peu de travaux se sont portés sur l'extraction des lois de comportements individuels directement au sein de grands groupes à partir d'expériences. En effet, l'occultation des individus par leurs congénères rend difficile l'extraction des données. Nous pouvons citer néanmoins le projet Starflag [6, 28, 29] sur les nuées d'étourneaux ainsi que les projets Pedigree sur les foules et Panurge sur les troupeaux de moutons que nous présentons ci-dessous. L'alternative consiste donc à rechercher les lois de comportements dans des petits groupes et à inférer des comportements dans les grands groupes. Cette démarche a été à la base des travaux [55, 94, 103].

Exposons maintenant les expériences menées dans les projets Panurge et Pedigree :

- Dans le cadre du projet Panurge, des expériences ont été menées sur l'initiation d'un mouvement collectif dans des petits groupes de moutons pendant une période de brout [102, 103]. L'expérience est la suivante : parmi un groupe de brebis placé dans une arène, l'une d'entre elles a été préalablement entraînée à rejoindre le bord de l'arène au déclenchement d'un signal sonore et visuel. Systématiquement, la brebis entrainée déclenche un mouvement collectif (voir Fig. 1) et ceci quel que soit le groupe considéré : le leadership est donc partagé par tous les membres du troupeau. La réactivité des autres brebis dites naïves à suivre la brebis initiatrice du mouvement est analysée en fonction de leur ordre de départ. Des expériences sur des grands groupes ont été aussi entreprises (voir Fig. 2). Elles sont en cours d'analyse.
- Le projet Pedigree vise à confronter différents modèles de déplacements de foules de piétons [65, 95] avec des données issues d'expériences. Une de ces expériences consiste à placer les piétons dans un anneau et à assigner à chacun d'eux un sens de

FIGURE 1 – 6 moutons dans une arène de 25 mètres de diamètre : suivi du mouton entraîné après déclenchement d'un signal visuel (© Marie-Hélène Pillot, Université Paul Sabatier, CRCA, Toulouse, France)

FIGURE 2 – 200 moutons dans un grand champ : observation d'un front de déplacement (© Marie-Hélène Pillot, Université Paul Sabatier, CRCA, Toulouse, France)

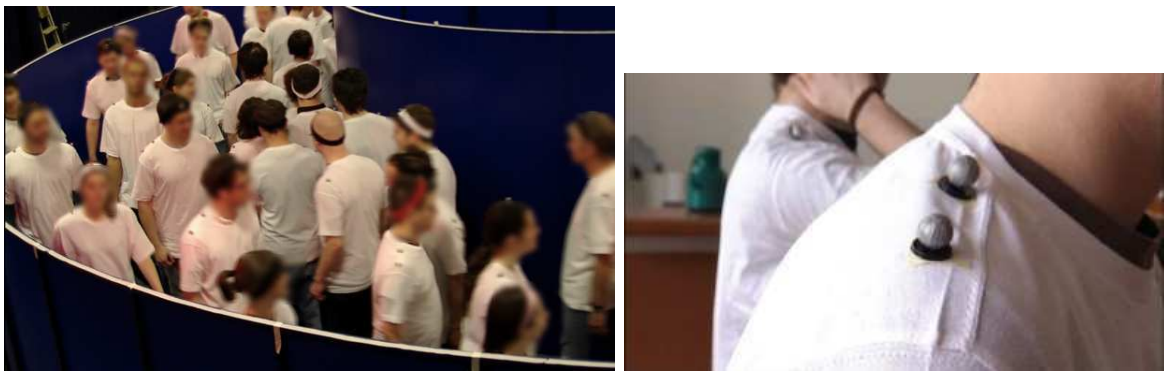


FIGURE 3 – Capture du mouvement de piétons dans un anneau : les piétons ayant le même sens de rotation se regroupent pour former des files (© Mehdi Moussaïd, Université Paul Sabatier, CRCA, Toulouse, France).

rotation (voir Fig. 3). Les déplacements sont enregistrés à l'aide de capteurs placés sur les épaules des individus. Des formations de files sont observées et l'analyse de leurs stabilités au cours du temps est à l'étude.

Dans la suite nous ne considérons que des modèles d'interactions minimalistes, au sens où ils ne sont pas extraits d'expériences et où leurs études sont vouées à explorer leurs propriétés qualitatives.

Dans les modèles que nous étudierons, les interactions entre les particules sont décrites par des équations de Newton avec des forces dites sociales. Considérant N particules et notant leurs positions et vitesses respectivement par $X_k \in \mathbb{R}^3$ et $V_k \in \mathbb{R}^3$ pour tout $k \in \{1, \dots, N\}$, les modèles s'écrivent :

$$\frac{dX_k}{dt} = V_k, \quad \frac{dV_k}{dt} = F_k. \tag{5}$$

Dans de nombreux travaux [2, 35], les forces F_k sont la somme de trois composantes : une force de répulsion à courte portée, une force d'alignement à moyenne portée et une force d'attraction à longue portée. Nous considèrerons pour le moment seulement le phénomène d'alignement et nous reportons la présentation des autres interactions au chapitre suivant.

La particularité de ces modèles par rapport à la physique dite classique tient au fait que ce sont des particules **auto-propulsées** [64] : les particules peuvent maintenir leur vitesse en puisant dans leurs ressources internes et donc la **quantité de mouvement n'est pas conservée** au cours du temps. La force F_k contient donc aussi en général un contrôle de la norme de la vitesse. Dans la suite nous considèrerons la **contrainte géométrique** suivante :

$$\forall k \in \{1, \dots, N\}, \quad |V_k| = 1. \tag{6}$$

1.3.2 Alignement : modèle de Vicsek.

D'un point de vue biologique, l'alignement correspond à de l'allélomimétisme : faire comme son semblable. Le modèle le plus simple est celui proposé par Vicsek [113] : c'est un modèle où le module des vitesses est constant et où chaque particule s'oriente selon

la moyenne des orientations de ses voisines. Ces interactions sont en compétition avec un bruit sur l'orientation.

Le modèle initial proposé dans [113] est en fait algorithmique : l'intensité des interactions est proportionnelle au pas de temps numérique. Dans le travail de Degond et Motsch [44], le modèle est transformé en un modèle continu en temps. Le modèle s'écrit alors :

$$\frac{dX_k}{dt} = V_k, \quad dV_k = (\mathrm{Id} - V_k \otimes V_k)\left(\nu \bar{V}_k dt + \sqrt{2d}\, dB_t^k\right), \tag{7}$$

où dB_t^k désigne un bruit blanc dans $\mathbb{R}^3$ et $\bar{V}_k$ l'orientation moyenne autour de la particule k :

$$\bar{V}_k = \frac{J_k}{|J_k|}, \quad J_k = \sum_{j=1}^{N} K(|X_j - X_k|)V_j, \tag{8}$$

où $K(|x-y|)$ désigne le noyau d'interaction et ne dépend que de la distance euclidienne entre deux particules x et y. Les paramètres ν et d sont respectivement liés à l'intensité de la force d'alignement et à l'intensité du bruit. La fonction $(\mathrm{Id} - V_k \otimes V_k)$ est la projection sur le plan orthogonal à V_k. Pour un vecteur $A \in \mathbb{R}^3$, nous avons :

$$(\mathrm{Id} - V_k \otimes V_k)A = A - (V_k \cdot A)V_k.$$

Cette fonction assure que l'accélération de la particule est orthogonale à la vitesse et par conséquent la norme de la vitesse est constante au cours du temps. La contrainte géométrique (6) est donc satisfaite.

Le modèle de Vicsek, version cinétique du modèle d'Ising utilisé pour décrire les matériaux ferromagnétiques, est un modèle a priori (non fondé sur l'expérience) dont le but est d'observer des **transitions de phase**, liées aux échelles de corrélation au sein du système. Pour les étudier, le paramètre d'ordre suivant est introduit :

$$\varphi = \frac{1}{N}\left|\sum_{i=1}^{N} V_i\right|.$$

Il mesure l'alignement global des particules. Ses variations en fonction de l'amplitude du bruit gaussien ou de la densité de particules permettent de définir la nature de la transition de phase. Le caractère continu ou discontinu de cette transition de phase a fait l'objet de vives recherches par les équipes de Vicsek et de Chaté [61, 113]. Par ailleurs, une étude de ces transitions sur une dynamique de Vicsek sur réseau a été menée par [1]. A titre de comparaison, l'observation d'une transition de phase a été mesurée dans une population de criquet par Couzin et al. [23].

Variantes du modèle. Plusieurs variantes du modèle ont fait l'objet d'études :

1. Le bruit peut être introduit non pas sur l'angle mais sur l'estimation de la vitesse moyenne J_k. Le bruit est dit vectoriel et du fait de la renormalisation, ce bruit est alors plus faible lorsque les particules sont alignées. Cette version est étudiée dans [57, 61].
2. Le modèle de Cucker-Smale [38, 39, 105] est obtenu en s'affranchissant de la contrainte sur le module de la vitesse et du bruit. Le comportement asymptotique de ce système est étudié dans [26, 62, 63].

3. Le modèle de Vicsek est un modèle de type champ-moyen. Une version collisionnelle est proposée dans [14].
4. Dans [52], un noyau d'interaction K non isotrope est considéré : K dépend aussi de V_k. De plus, la fréquence d'interaction d'alignement ν dépend de la densité locale.

1.3.3 Modèle macroscopique du modèle de Vicsek

Modèle cinétique et changement d'échelle hydrodynamique. L'objectif est maintenant d'obtenir un modèle à grande échelle en temps et en espace. Pour cela, nous poursuivons la même démarche que celle employée pour dériver les équations de la mécanique des fluides à partir des équations de Newton. Introduisant la fonction de distribution $f(x,v,t)$ dans l'espace des phases $x \in \mathbb{R}^3$, $v \in \mathbb{S}^2 = \{v,\ |v|=1\}$, $t \in \mathbb{R}^+$, le modèle cinétique champ-moyen associé à la dynamique particulaire (7)-(8) est le suivant :

$$\partial_t f + v\cdot\nabla_x f = Q(f), \tag{9}$$
$$Q(f) = -\nabla_v\cdot(Ff) + d\Delta_v f, \tag{10}$$
$$F(x,v,t) = \nu(\mathrm{Id} - v\otimes v)\bar{V}(x,t), \tag{11}$$
$$\bar{V}(x,t) = \frac{J(x,t)}{|J(x,t)|}, \quad J(x,t) = \int_{\mathbb{R}^3\times\mathbb{S}^2} K(|y-x|)wf(y,w,t)dydw. \tag{12}$$

L'opérateur $Q(f)$, modélisant les interactions de type Vicsek, est un modèle de type Fokker-Planck. Lorsque le bruit est nul ($d=0$), ce modèle peut être formellement dérivé de la dynamique particulaire [44]. Une modèle similaire a été obtenu pour le modèle de Cucker-Smale [63] via une hiérarchie de type BBGKY. Avec l'ajout du bruit ($d>0$), la limite champ-moyen nécessiterait une étude de la **propagation du chaos** [106, 114], c'est-à-dire le contrôle des corrélations dans le temps, et les interactions de type Vicsek n'étant pas sommes d'interactions binaires, la propagation du chaos est une question ouverte dans ce cas. Nous considérons donc dans la suite le modèle champ-moyen précédent et nous serons amenés à vérifier son bien-fondé numériquement.

Le changement d'échelle hydrodynamique consiste alors à faire le changement de variable suivant : $t' = \varepsilon t$, $x' = \varepsilon x$. Dans ces nouvelles variables, le système devient :

$$\begin{aligned}
&\partial_t f^\varepsilon + v\cdot\nabla_x f^\varepsilon = \frac{1}{\varepsilon}Q^\varepsilon(f^\varepsilon) + O(\varepsilon),\\
&Q^\varepsilon(f^\varepsilon) = -\nabla_v\cdot(F^\varepsilon f^\varepsilon) + d\Delta_v f^\varepsilon,\\
&F^\varepsilon(x,v,t) = \nu(\mathrm{Id} - v\otimes v)\Omega^\varepsilon(x,t),\\
&\Omega^\varepsilon(x,t) = \frac{j^\varepsilon(x,t)}{|j^\varepsilon(x,t)|}, \quad j^\varepsilon(x,t) = \int_{\mathbb{S}^2} wf^\varepsilon(x,w,t)dw,
\end{aligned} \tag{P_V^ε}$$

où nous pouvons remarquer que la force d'alignement F^ε est devenue locale. Le paramètre $1/\varepsilon$ peut être vu comme la fréquence des interactions adimensionnée.

Modèle macroscopique. Le modèle asymptotique (P_V^0) décrit l'évolution des quantités macroscopiques que sont la densité $\rho(x,t) = \int_{\mathbb{S}^2} f(x,v,t)dv$ et l'orientation moyenne

$\Omega(x,t)$:

$$\partial_t \rho + \nabla_x \cdot (c_1 \rho \Omega) = 0, \qquad (P_V^0)$$
$$\partial_t \Omega + c_2 \Omega \cdot \nabla_x \Omega = -\lambda (\mathrm{Id} - \Omega \otimes \Omega) \frac{\nabla_x \rho}{\rho},$$

où $\lambda = d/\nu$. Les coefficients c_1 et c_2 sont respectivement les vitesses de convection de la densité et de l'orientation et dépendent de λ. Du fait de la présence de l'opérateur $(\mathrm{Id} - \Omega \otimes \Omega)$ au second membre de l'équation sur Ω, nous retrouvons la non-conservativité de la quantité de mouvement déjà présente dans la dynamique particulaire.

La dérivation de ce modèle est présentée dans [44] et nous pouvons la résumer brièvement ainsi. Supposons que f^ε converge vers f alors le modèle (P_V^ε) implique que f est inclus dans le noyau de Q : $Q(f) = 0$. Or les fonctions appartenant au noyau de Q sont justement paramétrées par la densité et l'orientation moyenne. Pour obtenir la dynamique de ces paramètres, une méthode des moments est utilisée mais l'absence d'invariants collisionnels autre que le nombre de particules (du fait de la contrainte géométrique) nécessite de faire appel à des **invariants collisionnels généralisés**. Ces invariants collisionnels ne le sont en fait que pour les distributions avec une orientation prescrite. La robustesse de cette méthode a été confirmée dans [52]. Les corrections d'ordre $O(\varepsilon)$ de ce modèle sont présentées dans [46].

D'autres descriptions macroscopiques de la dynamique de Vicsek ont été proposées [14, 100]. La validation du modèle macroscopique présenté nécessite donc des simulations numériques (voir paragraphe 2.3 ci-dessous).

Dans le chapitre II, nous présenterons une version du modèle de Vicsek avec deux populations, des particules à l'arrêt et des particules en mouvement, ainsi que le modèle macroscopique associé. L'objectif est d'étudier le couplage entre deux modèles de Vicsek avec deux vitesses distinctes et l'impact que les termes d'échanges entre les deux populations ont sur la dynamique globale du système. Un résumé de ce travail se trouve au paragraphe 3.2.

1.4 Exemple 3 : Limite de congestion

1.4.1 Contexte : grégarisme et agrégation.

Les interactions dans les populations animales sont décrites par les forces d'alignement dont nous avons présenté les effets à grande échelle dans le paragraphe précédent mais aussi par les forces d'attraction et de répulsion. Celles-ci permettent notamment de garantir la cohésion d'un groupe. Les avantages et inconvénients biologiques de ce grégarisme sont multiples : facilité à échapper à un prédateur et à trouver un partenaire, compétition pour l'alimentation... Nous renvoyons à [97] pour de plus amples discussions et références sur ce sujet.

Modèles microscopiques. Reprenons le modèle individu-centré (5). Dans de nombreux modèles [2, 35, 67, 104], la force F_k est donc décomposée en termes attractifs à longue portée, en termes répulsifs à courte portée et en termes d'alignement à moyenne

portée. Il a été numériquement observé avec ces modèles des mouvements collectifs de rotation (on parle de moulin), des mouvements collectifs de translation,... suivant la valeur des paramètres du modèle [83]. Une grande partie de la littérature s'est donc attachée à comprendre le passage d'une forme à une autre : mentionnons des phénomènes d'hystérésis [35], des transitions de la translation au moulin sous l'effet du bruit [49],...

Dans cette thèse, nous nous intéressons tout particulièrement au problème de congestion et donc au traitement de la force de répulsion. Les particules ne pouvant se chevaucher, il existe une distance minimale $d_{\min}$ séparant chaque individu. Ceci induit l'existence d'une densité maximale ρ^*.

Distance entre particules. Parmi les modèles individus-centrés présents dans la littérature, un grand nombre ne prennent pas en compte cette distance minimale. La force de répulsion est même bien souvent bornée. La question est alors de savoir si une distance moyenne entre particules est conservée lorsque la taille de la population est augmentée ou bien s'il y a effondrement du système. La réponse est l'objet de l'étude des travaux [48, 93], où attraction et répulsion sont décrits par le potentiel de Morse.

Une première voie pour encoder la contrainte de distance minimale est de considérer une force de répulsion singulière en $d_{\min} > 0$: nous pouvons citer les travaux de Chaté et al. [57, 61], où le modèle de Vicsek est couplé avec des interactions de type Lennard-Jones. Signalons aussi une version du modèle de trafic routier Follow-the-Leader [12]. Une deuxième approche de type dynamique de contact a été proposée dans [90, 91], où les vitesses des particules sont projetées sur l'espace des vitesses permettant de préserver la distance minimale entre les particules.

Modèles macroscopiques et concentration. Pour ce qui concerne les modèles macroscopiques d'agrégation de particules, ils ne prennent en général pas en compte l'existence de densité maximale. Ce sont souvent des modèles du premier ordre (où les effets inertiels sont négligés) avec interaction non locale. Les études se sont tout d'abord intéressées à donner la bonne description des phénomènes d'agrégation en une dimension [92] puis en deux dimensions [111], sans ou avec diffusion [112]. Sans répulsion, des phénomènes de concentration ponctuelle ont lieu. Des études plus analytiques ont donc été menées pour donner des critères sur l'explosion en temps fini ou l'existence en tout temps de solution suivant la régularité du noyau d'interaction [15, 16].

Le même phénomène d'explosion en temps fini en l'absence de terme répulsif a lieu pour les modèles continus type Keller-Segel décrivant le déplacement de bactéries suivant le gradient de densité d'un chimio-attractant. Ce phénomène de concentration ponctuelle peut toutefois être évité avec une diffusion non linéaire [24, 79]. Un phénomène de concentration sur des réseaux unidimensionnels a été observé avec un modèle de type hyperbolique [50].

1.4.2 Contrainte de densité maximale

Contrairement à la plupart des modèles macroscopiques mentionnés précédemment, celui étudié dans cette thèse (chapitre IV) contient une contrainte sur la densité : $\rho < \rho^*$. Ce modèle est dérivé d'un modèle particulaire avec force de répulsion singulière, où les vitesses des particules sont de norme 1 comme pour le modèle de Vicsek (voir

paragraphe 1.3.2). C'est un modèle bidimensionnel. Il décrit la dynamique de la densité $\rho(x,t)$ et de l'orientation $\Omega(x,t) \in \mathbb{S}^1$, avec $x \in \mathbb{R}^2$ et $t > 0$:

$$\begin{aligned} &\partial_t \rho + \nabla_x \cdot (\rho\Omega) = 0, \\ &\partial_t \rho\Omega + \nabla_x \cdot (\rho\Omega \otimes \Omega) = -(\mathrm{Id} - \Omega \otimes \Omega)\,(\varepsilon \nabla_x p(\rho))\,, \end{aligned} \tag{P_C^ε}$$

où p modélise une **pression singulière** en ρ^* :

$$p(\rho) \xrightarrow[\rho \to \rho^*]{} +\infty.$$

Le paramètre ε mesure pour ce problème la singularité de la force de répulsion $\varepsilon \nabla_x p(\rho)$. La limite de ce problème lorsque $\varepsilon \to 0$ s'écrit ainsi :

$$\begin{aligned} &(\rho^* - \rho)\bar{p} = 0, \\ &\partial_t \rho + \nabla_x \cdot (\rho\Omega) = 0, \\ &\partial_t \rho\Omega + \nabla_x \cdot (\rho\Omega \otimes \Omega) = -(\mathrm{Id} - \Omega \otimes \Omega)\,(\nabla_x \bar{p})\,, \\ &|\Omega| = 1, \end{aligned} \tag{P_C^0}$$

où $\bar{p}$ devient une nouvelle inconnue de notre problème. Elle est non nulle seulement dans les domaines de densité maximale.

La contrainte de densité maximale est aussi présente dans les modèles pour les fluides bi-phasiques (par exemple pour les mélanges gaz-liquide) où ρ joue alors le rôle de la concentration d'un des deux constituants. Nous renvoyons à [68, 109] pour une présentation de ces modèles. Dans [20], un modèle asymptotique unidimensionnel similaire à (P_C^0) est proposé (sans la contrainte géométrique sur la vitesse du fluide). Il est obtenu comme une limite de relaxation d'un modèle biphasique lorsque le ratio des densités des deux fluides en présence devient faible. L'existence de solutions faibles a été démontrée dans [11] en exhibant des solutions nommées bouchons collants, c'est-à-dire des domaines de concentration maximale. Un modèle similaire avec diffusion a été étudié dans [85].

L'asymptotique de congestion portée par la pression singulière a été à l'origine développée pour la modélisation des bouchons dans les trafics routiers [12, 13] à partir du modèle d'Aw-Rascle [4, 5]. Pour ces modèles, la vitesse est unidirectionnelle et n'est pas contrainte. Sur le même principe, un modèle bidimensionnel pour les foules a été proposé par Bellomo et Dogbé [8], mais dans ce cas la force répulsive agit seulement dans une direction fixée a priori (dans ce cas, la sortie de secours). Enfin, la contrainte de densité maximale a été aussi prise en compte dans le cadre de modèles du premier ordre pour les trafics [31, 84].

1.4.3 Problème asymptotique : transition compressible/incompressible.

Le modèle (P_C^0) a une structure particulièrement riche : dans les domaines de densité maximale $\rho = \rho^*$, le fluide se comporte comme un fluide *incompressible* tandis que dans les domaines de densité plus faible $\rho < \rho^*$, le fluide est *compressible*. Dans ces zones compressibles où $\rho < \rho^*$, le modèle s'apparente au modèle de gaz sans pression développé notamment pour décrire les phénomènes de formation de galaxie [116] et qui conduirait sans la contrainte de densité à des phénomènes de concentration ponctuelle. Pour ce qui concerne les propriétés mathématiques de ce modèle de gaz sans pression, nous renvoyons aux travaux suivants [19, 22, 59, 115] .

Analyse du problème asymptotique. Le problème (P_C^0) est incomplet car pour définir une solution, des conditions aux interfaces entre les domaines compressibles et incompressibles doivent être fournies. Une partie de ces conditions sont obtenues en considérant la limite $\varepsilon \to 0$ des solutions du **problème de Riemann** unidimensionnel. Une étude similaire a été menée pour le modèle de trafics routiers [12], trois différences importantes peuvent toutefois être soulignées :

1. Le modèle (P_C^0) est non-conservatif, donc il a fallu faire le choix d'une formulation conservative.
2. Le modèle étudié n'est pas un système de Temple (pour lequel courbes de raréfaction et courbes de choc coïncident) et les courbes de choc sont définies implicitement. L'étude, bien que similaire, en est plus technique.
3. Le problème de Riemann entre deux états congestionnés est cette fois mal posé. Une solution en termes de bouchons collants a été donnée pour combler en partie ce déficit.

Ainsi, nous avons déterminé les principales caractéristiques des interfaces entre les régions compressibles et les régions incompressibles mais des aspects plus géométriques tels que la collision bidimensionnelle entre deux régions de densité maximale nous font défaut. Un résumé plus détaillé de ce travail se trouve au paragraphe 3.4.

2 Schémas numériques préservant l'asymptotique

2.1 Présentation des méthodes

Les méthodes préservant l'asymptotique sont des méthodes numériques permettant de résoudre un problème (P^ε) dépendant d'un paramètre ε et de capturer sa limite.

Pour introduire le problème, considérons l'équation différentielle raide suivante :

$$y_\varepsilon(0) = a \in \mathbb{R}, \quad \frac{dy_\varepsilon}{dt} = -\frac{1}{\varepsilon} y_\varepsilon, \tag{P_E^ε}$$

et le problème asymptotique associé :

$$y_0(0) = a \in \mathbb{R}, \quad y_0(t) = 0, \ \forall t > 0. \tag{P_E^0}$$

Notant $y_\varepsilon^n = y_\varepsilon(n\Delta t)$, $n \in \mathbb{N}$, les deux schémas numériques suivants sont consistants avec le problème P_E^ε et inconditionnellement stables :

$$\text{méthode des trapèzes,} \qquad y_\varepsilon^{n+1} = \frac{(1 - \Delta t/(2\varepsilon))}{(1 + \Delta t/(2\varepsilon))} y_\varepsilon^n,$$

$$\text{méthode d'Euler implicite,} \qquad y_\varepsilon^{n+1} = \frac{1}{(1 + \Delta t/\varepsilon)} y_\varepsilon^n.$$

Pourtant, en prenant la limite $\varepsilon \to 0$ dans les schémas numériques, seul le schéma d'Euler implicite fournit une discrétisation de P_E^0 : il est inconditionnellement consistant [1]. Donnons à présent la définition d'un schéma préservant l'asymptotique.

1. On parle aussi de L-stabilité pour les équations différentielles.

Schéma Préservant l'Asymptotique (AP). Soit (P^ε) un problème dépendant d'un paramètre ε et $(P^\varepsilon_{\Delta t,\Delta x})$ un schéma numérique associé, où Δt et Δx désignent les pas de temps et d'espace. Ce schéma numérique est dit AP si à Δt et Δx fixés, le schéma numérique limite $(P^0_{\Delta t,\Delta x}) = \lim_{\varepsilon\to 0}(P^\varepsilon_{\Delta t,\Delta x})$ fournit un schéma numérique consistant avec le problème limite $(P^0) = \lim_{\varepsilon\to 0}(P^\varepsilon)$. Cette propriété est décrite par le diagramme commutatif suivant :

$$\begin{array}{ccc} P^\varepsilon_{\Delta t,\Delta x} & \xrightarrow{\varepsilon\to 0} & P^0_{\Delta t,\Delta x} \\ {\scriptstyle \Delta t,\Delta x\to 0}\downarrow & & \vdots\,{\scriptstyle \Delta t,\Delta x\to 0} \\ P^\varepsilon & \xrightarrow[\varepsilon\to 0]{} & P^0 \end{array}$$

L'appellation "asymptotic preserving scheme" a été introduite par S. Jin dans [69].

Il faut en outre s'assurer que le schéma soit **uniformément stable** par rapport à ε pour que le schéma soit praticable. Par exemple pour des problèmes de transport, alors que la stabilité des schémas numériques classiques nécessiterait la résolution d'une condition de type CFL $\Delta t = O(\varepsilon\Delta x)$, l'uniforme stabilité se traduit par une condition indépendante de ε : $\Delta t = O(\Delta x)$.

L'uniforme stabilité permet donc d'envisager des simulations numériques **sous-résolues** en temps et en espace, $\Delta t, \Delta x \gg \varepsilon$, tandis que le caractère AP permet de garantir que la solution est bien une solution approchée du problème.

Donnons à présent les principaux **ingrédients** pour construire un schéma numérique préservant l'asymptotique :

1. **Réécriture de la limite singulière.** L'intérêt et la difficulté pour trouver des schémas préservant l'asymptotique réside dans le fait que le problème (P^ε) et sa limite (P^0) peuvent être de natures mathématiques différentes. Pour construire un schéma préservant l'asymptotique, il faut donc trouver une formulation commune entre le problème et sa limite pour laquelle la limite $\varepsilon \to 0$ est "régularisée" : le problème et sa limite sont sous la même forme mathématique, ce qui permet d'envisager le même traitement numérique. C'est ce que nous avons déjà entrepris pour la limite quasi-neutre dans les plasmas (voir paragraphe 1.2).
2. **Implicitation des termes raides.** Ceci garantit l'absolue stabilité du schéma. Cette étape peut s'effectuer à partir du schéma semi-discrétisé en temps. Toutefois, il faut garder en mémoire que l'implicitation doit être effective : si par exemple, l'implicitation est assurée par un algorithme de point fixe, alors il faut s'assurer que le nombre maximum d'étapes utilisé soit indépendant de ε.
3. **Choix d'un schéma en espace "bien adapté".** La discrétisation spatiale devra par exemple satisfaire des propriétés de conservativité, des critères d'entropie... suivant les problèmes considérés.

Cette démarche a été explorée dans plusieurs contextes :

- Les schémas de relaxation pour les systèmes hyperboliques non-linéaires [73]. Les problèmes hyperboliques non-linéaires sont réécrits en des systèmes hyperboliques *linéaires* avec terme source raide. Une méthode de splitting en temps est combinée avec un traitement implicite du terme raide. Le problème asymptotique est ici artificiellement introduit comme un outil numérique.

– La limite hydrodynamique ou de diffusion dans les équations cinétiques. Le traitement numérique est là aussi essentiellement basé sur un splitting entre la partie transport et la partie collisionnelle de l'équation et un traitement implicite de cette dernière. Pour la limite de diffusion, un splitting naïf conduit à des schémas ne préservant pas l'asymptotique [72], il faut donc trouver une reformulation adaptée. Pour des équations discrètes en espace, le lecteur pourra consulter [69, 72]. La difficulté dans ces méthodes est de traiter implicitement le terme de collision. Pour cela, un grand nombre de méthodes ont été développées en s'appuyant sur une décomposition de Chapman-Enskog [75, 76], ou plus généralement sur un développement en série [32, 53, 96], ou bien en décomposant la fonction de distribution en vitesse en composante paire et impaire [70, 71], ou encore via une décomposition équilibre/hors équilibre [9, 10, 27, 82].

Nous pouvons citer enfin la limite de diffusion de l'équation de Ginzburg-Landau [42]. Nous présenterons dans la suite d'autres méthodes dans le cas de la limite quasi-neutre [7, 36, 37] et dans la limite bas-Mach pour les équations d'Euler [43, 45].

Enfin, citons plusieurs travaux où les preuves de la consistance et/ou de la stabilité uniforme ont été abordées : [7, 36, 56, 77, 87].

Avantages et inconvénients. Quels sont les avantages d'un schéma préservant l'asymptotique ? Tout d'abord, l'uniforme stabilité des schémas numériques permet de ne pas résoudre les petites échelles et donc de réduire le coût numérique. Le deuxième avantage est que nous bénéficions d'une même méthode pour le problème et sa limite. Nous n'avons donc pas besoin d'utiliser deux méthodes numériques distinctes, l'une valide pour les domaines où $\varepsilon \ll 1$ et l'autre pour les domaines où $\varepsilon = O(1)$. La gestion numérique de l'interface entre ces domaines nécessiterait la recherche des conditions aux limites adéquates ainsi que le traitement numérique de la position de cette interface (par des méthodes ALE, des méthodes de Front-Tracking, des méthodes Volume-Of-Fluid, des méthodes Level-set, des méthodes d'interface diffuse).

Toutefois, l'usage de grands pas de temps et d'espace accroît la diffusion numérique. Il faut donc recourir à des méthodes de décomposition de domaine si la précision n'est pas satisfaisante.

2.2 Méthode numérique particulaire AP pour la limite quasi-neutre

Rappelons le système de Vlasov-Poisson modifié introduit au paragraphe 1.2 :

$$\partial_t f + v \cdot \nabla_x f + \nabla_x \phi \cdot \nabla_v f = 0, \tag{13}$$

$$-\nabla_x \cdot \left(\left(1 + \rho + \varepsilon^2 \partial_{tt}^2 \right) \nabla_x \phi \right) = -\nabla_x^2 : S, \tag{14}$$

L'introduction de ce problème modifié va permettre de développer des méthodes AP. En effet, contrairement au système initial, ce système contient une équation pour le potentiel ϕ qui est du même type pour le système perturbé avec $\varepsilon > 0$ ou pour le système limite avec $\varepsilon = 0$. Avant de présenter plus en détails la méthode, introduisons la méthode numérique employée, à savoir les méthodes particulaires.

2.2.1 Méthodes particulaires

Les méthodes particulaires consistent à résoudre les équations des caractéristiques de l'équation de Vlasov (13). Considérons la distribution empirique dans l'espace des phases :

$$f_0^N(x,v) = \frac{1}{N}\sum_{j=1}^{N} \delta(x - X_i^0) \otimes \delta(v - V_j^0), \quad \text{avec } \forall j \in \{1,\dots,N\},\ (X_j^0, V_j^0) \in \mathbb{R}^d,$$

où δ désigne la distribution de Dirac et $\otimes$ le produit tensoriel. Alors la distribution

$$f^N(x,v,t) = \frac{1}{N}\sum_{j=1}^{N} \delta(x - X_j(t)) \otimes \delta(v - V_j(t)),$$

où $\{(X_j, V_j),\ j \in \{1,\dots,N\}\}$ sont solutions des équations des caractéristiques de (13) :

$$\begin{aligned} &\frac{dX_j}{dt} = V_i, \quad \frac{dV_j}{dt} = \nabla_x \phi(X_j), \\ &X_j(0) = X_j^0, \quad V_j(0) = V_j^0, \end{aligned} \tag{15}$$

est solution de l'équation de Vlasov avec pour distribution initiale f_0^N. De plus une propriété de stabilité implique que si f_0^N converge vers f_0 quand $N \to +\infty$, alors f^N converge vers la solution de l'équation de Vlasov avec condition initiale f_0 [108]. Les méthodes particulaires consistent donc à distribuer N particules suivant la loi f_0 et à résoudre les équations des caractéristiques. Notons que ces particules numériques suivent les mêmes lois dynamiques que les particules physiques.

Le potentiel électrostatique ϕ, présent dans les équations caractéristiques, est obtenu en résolvant l'équation de Poisson (14) par l'intermédiaire d'un maillage en espace. Il faut donc définir les quantités macroscopiques associées à la distribution de particules (la densité ρ et le flux de quantité de mouvement S) sur le maillage : cette étage est appelée **assignation**. Ensuite, une fois le calcul du potentiel effectué, il faut l'**interpoler** sur la position des particules pour calculer le déplacement des particules. La méthode d'assignation peut être vue comme une méthode de régularisation de la fonction de distribution f^N. Considérant une fonction $h \in \mathcal{C}(\mathbb{R}^d)$ à support compact telle que $\int_{\mathbb{R}^d} h(x)dx = 1$ et ses dilatations associées $h_\alpha(x) = \frac{1}{\alpha^d} h\left(\frac{x}{\alpha}\right)$, l'assignation de la densité se fait ainsi :

$$\tilde{\rho}^N(x,t) = \rho^N(.,t) *_x h_{\Delta x} := \int_{\mathbb{R}^d} \rho^N(y,t) h_{\Delta x}(y-x)dy,$$

où Δx désigne le pas d'espace et $\rho^N(x,t)$ la densité associée à $f^N(x,v,t)$:

$$\rho^N(x,t) = \int_{\mathbb{R}^d} f^N(x,v,t)dv = \frac{1}{N}\sum_{j=1}^{N} \delta(x - X_j(t)).$$

Lorsque la fonction de régularisation h est la fonction chapeau :

$$h(x) = \begin{cases} 1 - |x| & \text{si } x \in B(0,1) = \left\{x \in \mathbb{R}^d, |x| \leqslant 1\right\}, \\ 0 & \text{sinon,} \end{cases}$$

la méthode est appelée méthode Particle-In-Cell (PIC). L'étape d'interpolation utilise le même procédé de régularisation.

Pour une présentation plus générale des méthodes particulaires, nous renvoyons aux ouvrages de Birdsall et Langdon [17] et d'Hockney et Eastwood [66]. La convergence d'une telle méthode a été étudiée dans le cas unidimensionnel périodique par [33] et en dimensions supérieures par [54], lorsque les particules sont initialement sur un réseau de l'espace des phases.

Autres méthodes. Les méthodes particulaires sont très utilisées dans la pratique et ceci bien qu'elles soient très bruitées. De plus, lorsque les particules sont distribuées initialement aléatoirement suivant la fonction de distribution f_0, la convergence de la méthode n'est seulement qu'en $O(1/\sqrt{N})$, où N désigne le nombre de particules c'est-à-dire la taille des données. Deux autres méthodes ont donc été développées pour obtenir des méthodes plus précises : la méthode semi-lagrangienne [107], la méthode de type volume fini (voir [51] et référence interne). Elles reposent toutes deux sur un maillage en espace et aussi en vitesse et donc nécessitent des coûts plus élevés en termes de mémoire.

2.2.2 Limite quasi-neutre et méthode numérique AP.

Nous avons présenté au paragraphe 1.2 le problème de la quasi-neutralité dans les plasmas. D'un point de vue numérique, en plus de la condition de stabilité du type CFL $\Delta t \leqslant v_{\max}\Delta x$, où $v_{\max}$ désigne la vitesse maximale des particules, les méthodes numériques classiques doivent résoudre les petites échelles :

$$\Delta t \leqslant \varepsilon, \quad \Delta x \leqslant \varepsilon.$$

Des méthodes ont été développées pour s'affranchir de ces deux contraintes. Elles consistent à impliciter le potentiel dans le calcul des trajectoires des particules : la méthode Moment-Implicit [89] calcule ce potentiel implicite à partir de l'équation de Poisson et d'une densité implicite calculée à partir de l'équation de conservation de la masse, la méthode Direct-Implicit [80] calcule ce potentiel par prédiction-correction via le déplacement des particules.

Notre méthode, développée au chapitre I, consiste elle aussi à impliciter le calcul du potentiel mais elle s'appuie sur l'équation de Poisson reformulée (14) et non sur l'équation de Poisson. La discrétisation de l'équation de Poisson reformulée est obtenue par le même mode opératoire que celui utilisé pour obtenir l'équation de Poisson reformulée : il suffit d'insérer l'équation sur la quantité de mouvement dans l'équation de conservation de la masse et d'utiliser l'équation de Poisson pour obtenir une équation sur le potentiel. Plusieurs cas tests ont été à l'étude : ils ont montré le caractère asymptotique de notre schéma. Un résumé de ce travail se trouve au paragraphe 3.1.

Cette technique avait déjà été utilisée pour résoudre le système d'Euler-Poisson [37] avec des schémas de type volume fini. Une analyse de stabilité linéaire est faite dans [36]. La même méthodologie a aussi été entreprise dans le cadre des méthodes semi-lagrangiennes [7].

2.3 Validation numérique du modèle de Vicsek macroscopique

Le modèle de Vicsek macroscopique (P_V^0) étant **non-conservatif**, le problème est **sous-déterminé** au sens où il faut donner une signification mathématique au terme non

conservatif $(\mathrm{Id} - \Omega \otimes \Omega)\nabla_x \rho$. Pour illustrer ce phénomène, considérons le système de Riemann suivant :

$$\begin{aligned} &\partial_t u + g(u)\partial_x u = 0, \\ &u(0,x) = \begin{cases} u_\ell, & \text{si } x \in]-\infty, 0], \\ u_r, & \text{si } x \in]0, +\infty[, \end{cases} \end{aligned}$$

avec $u_\ell, u_r \in \mathbb{R}^d$ et $g : \mathbb{R}^d \to \mathcal{M}_d$, où $\mathcal{M}_d$ désigne l'ensemble des matrices de taille d. Si u est discontinue, alors nous devons définir le produit d'une fonction discontinue $g(u)$ avec une distribution $\partial_x u$. Par l'intermédiaire d'un chemin $\phi : [0,1] \to \mathbb{R}^d$ reliant u_g à u_d ($\phi(0) = u_g$, $\phi(1) = u_d$), nous pouvons définir une régularisation de u. Nous constatons alors que la définition qui résulte de cette régularisation dépend du chemin choisi ϕ :

$$g(u)\partial_x u = \lim_{\delta \to 0} \left[g(u^\delta)\partial_x u^\delta \right]_\phi = \left(\int_0^1 g(\phi(s))\phi'(s) ds \right) \delta_0.$$

Les solutions en ondes de choc sont donc dépendantes des chemins utilisés. Dans [81], les chemins proposés sont ceux associés aux profils de chocs visqueux, résultant d'une régularisation parabolique du problème. Des méthodes numériques ont été développées à partir de cette définition [99].

Concernant notre problème (P_V^0), la question est alors de savoir quel est le critère pour sélectionner la bonne solution, c'est-à-dire celle qui correspond aux comportements macroscopiques de la dynamique particulaire. Contrairement au problème de physique classique, nous ne disposons pas a priori de notion d'entropie qui permette cette sélection. C'est numériquement que nous avons effectué cette étude dans le chapitre III : nous proposons différents schémas numériques qui a priori définissent différentes notions de solutions du modèle macroscopique et nous les comparons à la simulation du système particulaire. Cette étude est résumée au paragraphe 3.3.

Lien avec les schémas préservant l'asymptotique. Un de ces schémas est notamment basé sur la relaxation de la contrainte et c'est d'ailleurs celui qui s'avère capturer le mieux la dynamique particulaire. Nous considérons le système suivant :

$$\begin{aligned} &\partial_t \rho + \nabla_x \cdot (\rho\Omega) = 0, \\ &\partial_t (\rho\Omega) + c\nabla_x \cdot (\rho\Omega \otimes \Omega) + \frac{\nu}{d} \nabla_x \rho = \frac{\rho}{\varepsilon}(1 - |\Omega|^2)\Omega, \end{aligned}$$

où ε est ici un outil numérique caractérisant l'échelle de temps durant laquelle a lieu la relaxation. Une méthode de splitting est ensuite utilisée : à chaque pas de temps, le système évolue sans la contrainte puis la vitesse obtenue est normalisée. En prenant la limite $\varepsilon \to 0$ dans ce modèle, nous obtenons formellement le modèle (P_V^0).

2.4 Méthode numérique AP pour la limite de congestion

Au paragraphe 1.4, nous avons introduit la limite de congestion dans le cadre des équations d'Euler avec une contrainte de densité maximale et une contrainte géométrique sur la vitesse du fluide ($u \in \mathbb{S}^1$). Pour ce qui concerne la contrainte géométrique, nous

avons déjà mentionné au paragraphe précédent qu'une méthode numérique valide (au moins pour le modèle de Vicsek) consiste à effectuer une méthode de splitting. Nous nous intéressons donc ici à l'obtention numérique des solutions sans la contrainte géométrique.

2.4.1 Méthode numérique pour la limite bas-Mach.

Pour introduire le problème, considérons le système d'Euler isentropique adimensionné satisfait par la densité $\rho(x,t)$ et la quantité de mouvement $q(x,t)$, avec $x \in \mathbb{R}^2$ et $t \in \mathbb{R}^+$, introduit au paragraphe 1.1 :

$$\begin{aligned} &\partial_t \rho + \nabla_x \cdot q = 0, \\ &\partial_t q + \nabla_x \cdot \left(\frac{q \otimes q}{\rho} \right) + \frac{1}{\varepsilon^2} \nabla_x p(\rho) = 0, \\ &p(\rho) = \rho^\gamma, \end{aligned} \tag{P^ε_{BM}}$$

où γ est une constante physique et le paramètre ε désigne le nombre de Mach. Le problème limite est le système d'Euler incompressible satisfait par la densité ρ et la vitesse $u = q/\rho$:

$$\begin{aligned} &\rho = \rho_0, \quad \nabla_x \cdot u = 0, \\ &\partial_t u + \nabla_x \cdot (u \otimes u) + \nabla_x \bar{p} = 0, \end{aligned} \tag{P^0_{BM}}$$

où ρ_0 est une constante et $\bar{p}$ est le multiplicateur de Lagrange de la contrainte d'incompressibilité : $\nabla_x \cdot u = 0$. Le système (P^ε_{BM}) génère des ondes de vitesses :

$$\lambda = u \pm \frac{1}{\varepsilon} \sqrt{p'(\rho)}.$$

Ceci rend impossible l'utilisation de schémas classiques pour des valeurs de ε petites car ces schémas doivent satisfaire la condition CFL : $\Delta t = O(\lambda \Delta x)$. Les méthodes préservant l'asymptotique pour ce problème [43, 45, 78] sont basées sur les deux ingrédients suivants :

1. La pression satisfait l'équation elliptique suivante :

$$\varepsilon^2 \partial_{t^2} \rho - \varepsilon^2 \nabla_x^2 : \left(\frac{q \otimes q}{\rho} \right) - \Delta p(\rho) = 0. \tag{16}$$

   Cette équation pouvant être résolue implicitement, elle permet d'impliciter le terme raide dans l'équation sur la quantité de mouvement et donc d'obtenir une stabilité asymptotique du schéma. De plus, elle ne dégénère pas quand $\varepsilon \to 0$.

2. Un schéma numérique volume fini conservatif et entropique pour le système d'Euler permet de capturer la solution entropique. La discrétisation spatiale de l'équation elliptique (16) doit donc être en conformité avec un tel schéma.

Dans la pratique, pour construire le schéma numérique, il suffit donc de considérer la semi-discrétisation en temps suivante :

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^{n+1} = 0, \tag{17}$$

$$\frac{q^{n+1} - q^n}{\Delta t} + \nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} \right) + \nabla_x p(\rho^{n+1}) = 0, \tag{18}$$

puis d'introduire l'expression de q^{n+1} donnée par (18) dans l'équation sur la densité (17) :

$$\frac{\rho^{n+1}-\rho^n}{\Delta t}+\nabla_x\cdot q^n-\Delta t\nabla_x^2:\left(\frac{q^n\otimes q^n}{\rho^n}\right)+\Delta t\Delta_x p(\rho^{n+1})=0,$$
$$\frac{q^{n+1}-q^n}{\Delta t}+\nabla_x\cdot\left(\frac{q^n\otimes q^n}{\rho^n}\right)+\nabla_x p(\rho^{n+1})=0.$$

Pour obtenir le schéma numérique discrétisé en temps et en espace, il faut réaliser les mêmes opérations en partant des schémas conservatifs discrétisés en espace.

2.4.2 Méthode numérique avec contrainte de densité maximale

Dans le chapitre V, nous appliquons la démarche expliquée ci-dessus au système d'Euler contraint :

$$\begin{aligned}&\partial_t\rho+\nabla_x\cdot q=0,\\&\partial_t q+\nabla_x\cdot\left(\frac{q\otimes q}{\rho}\right)+\nabla_x\varepsilon p(\rho)=0,\end{aligned}\qquad(P^{\varepsilon}_{EC})$$

avec une pression singulière : $p(\rho)$ tend vers $+\infty$ quand la densité ρ tend vers la densité maximale ρ^*. Ce système est celui introduit au paragraphe 1.4.2 sans contrainte sur le module de la vitesse. Le modèle limite est :

$$\begin{aligned}&(\rho-\rho^*)\bar{p}=0,\\&\partial_t\rho+\nabla_x\cdot q=0,\\&\partial_t q+\nabla_x\cdot\left(\frac{q\otimes q}{\rho}\right)+\nabla_x\bar{p}=0.\end{aligned}\qquad(P^{0}_{EC})$$

La principale différence avec la limite bas-Mach est que (P^0_{EC}) est un modèle bi-fluide : dans les domaines congestionnés de densité maximale $\rho=\rho^*$, il décrit un fluide incompressible et dans les domaines non congestionnés de densité plus faible $\rho<\rho^*$, la pression est nulle et le modèle décrit un fluide fortement compressible. L'objectif est donc de capturer deux dynamiques distinctes ainsi que la dynamique des interfaces entre les domaines congestionnés et non-congestionnés.

Nous proposons deux schémas numériques, l'un directement issu de la méthodologie expliquée ci-dessus et l'autre utilisant de plus la décomposition de Hodge :

$$q=a-\nabla_x\phi,\quad \nabla_x\cdot a=0.$$

L'utilisation de la décomposition de Hodge pour la limite bas-Mach a été proposée dans [43]. Un résumé détaillé de l'étude de ces deux schémas se trouve au paragraphe 3.5.

3 Résumé des travaux présents dans cette thèse

3.1 Méthode Particle-In-Cell pour la limite quasi-neutre

Dans le chapitre I, nous présentons deux méthodes particulaires pour le système Vlasov-Poisson reformulé (13)-(14). Nous montrons numériquement qu'elles préservent

l'asymptotique quasi-neutre grâce à plusieurs cas tests unidimensionnels. Nous les comparons de plus à une autre méthode particulaire, la méthode Direct-Implicit [80], et à la méthode semi-lagrangienne [7].

Notant $(X)_j^m = X_j(m\Delta t)$ et $(V)_j^m = V_j(m\Delta t)$ la position et la vitesse de la j-ème particule numérique au temps $t^m = m\Delta t$, les particules évoluent suivant le schéma semi-implicite suivant :

$$\frac{(X)_j^{m+1} - (X)_j^m}{\Delta t} = (V)_j^{m+1}, \tag{19}$$

$$\frac{(V)_j^{m+1} - (V)_j^m}{\Delta t} = \nabla_x \phi^{m+1}((X)_j^m), \tag{20}$$

discrétisant les équations des caractéristiques (15) de l'équation de Vlasov. L'évaluation implicite du potentiel ϕ^{m+1} s'appuie sur des discrétisations de l'équation de Poisson reformulée (14). La première discrétisation est la suivante :

$$-\nabla_x \cdot \Big(((\Delta t)^2 n^m + \varepsilon^2) \nabla_x \phi^{m+1} \Big) = -(\Delta t)^2 \nabla_x^2 : S^m + 2(1 - n^m) - (1 - n^{m-1}), \tag{21}$$

où $n^m = n(m\Delta t)$ et $S^m = S(m\Delta t)$ sont la densité de particules et le flux convectif de la quantité de mouvement calculés à partir de la distribution des particules au temps t^m par une méthode d'assignation. Les discrétisations spatiales sont faites à l'aide des différences finies. Les équations (19)-(20)-(21) constituent la méthode PICAP-1. Remarquons que cette méthode nécessite la connaissance de la densité au temps t^{-1} pour le calcul du potentiel au temps t^1. Si celle-ci fait défaut, il faut recourir à un schéma explicite au premier pas de temps, ce qui peut engendrer des instabilités dans la limite $\varepsilon \to 0$. La seconde discrétisation de l'équation de Poisson reformulée (14) est la suivante :

$$-\nabla_x \cdot \Big(((\Delta t)^2 n^m + \varepsilon^2) \nabla_x \phi^{m+1} \Big) = -(\Delta t)^2 \nabla_x^2 : S^m + 1 - n^m + \Delta t (\nabla_x \cdot (nu)^m). \tag{22}$$

Cette discrétisation peut être obtenue directement en insérant une discrétisation de l'équation de conservation de la quantité de mouvement :

$$\frac{nu^{m+1} - nu^m}{\Delta t} + \nabla_x \cdot S^m = n^m \nabla_x \phi^{m+1},$$

dans la discrétisation de l'équation de conservation de la masse :

$$\frac{n^{m+1} - n^m}{\Delta t} + \nabla_x \cdot (nu^{m+1}) = 0,$$

et en utilisant l'équation de Poisson $n^{m+1} = \Delta \phi^{m+1} - 1$. Les équations (19)-(20)-(22) constituent la méthode PICAP-2.

Ces méthodes préservent l'asymptotique quasi-neutre car elles fournissent à la limite $\varepsilon \to 0$ une discrétisation consistante avec la limite $\varepsilon \to 0$ de l'équation de Poisson reformulée :

$$-\nabla_x \cdot \Big(n \nabla_x \phi \Big) = -\nabla_x^2 : S.$$

De plus, la stabilité des schémas est apportée par l'implicitation du potentiel dans (20).

Pour valider nos deux méthodes numériques, nous les avons comparées numériquement avec la méthode dite classique (les trajectoires des particules sont intégrées explicitement et le potentiel explicite est calculé avec une discrétisation de l'équation de Poisson) ainsi qu'avec diverses variantes de la méthode Direct-Implicit [80], qui est une méthode de type prédiction-correction. Plusieurs cas tests unidimensionnels ont été considérés.

Tous les cas tests ont montré que les méthodes PICAP sont stables lorsque les échelles de quasi-neutralité ne sont pas résolues, c'est-à-dire lorsque les pas d'espace et de temps vérifient $\Delta x > \varepsilon$, $\Delta t > \varepsilon$. La méthode classique est quant à elle instable.

Les deux premiers cas tests, un cas test stationnaire et une perturbation spatiale périodique, ont permis de mesurer la dissipation d'énergie des différentes méthodes particulaires. Dans le cas où les échelles de quasi-neutralité sont résolues, c'est-à-dire lorsque les pas d'espace et de temps vérifient $\Delta x < \varepsilon$, $\Delta t < \varepsilon$, l'énergie totale est fortement dissipée par les schémas PICAP. Cette dissipation d'énergie est moins importante pour les schémas classique et Direct Implicit. Toutefois, dans le cas sous-résolu où les pas de temps et d'espace vérifient $\Delta x > \varepsilon$, $\Delta t > \varepsilon$, la dissipation d'énergie des schémas PICAP et Direct Implicit devient très faible, notamment parce que le nombre de particules par maille est plus élevé et le nombre d'assignation-interpolation par unité de temps est réduit. Dans le cas test de perturbation périodique, l'utilisation de deux espèces de particules (ions et électrons) avec un ratio de masse réaliste démontre la robustesse des méthodes PICAP.

Une étude de l'amortissement Landau et de l'instabilité double faisceaux est ensuite effectuée. Les méthodes PICAP ont pu être comparées aux résultats théoriques et numériques de la méthode semi-lagrangienne [7]. Dans les cas résolus et sous-résolus, les méthodes PICAP donnent des résultats stables mais moins précis que ceux fournis par la méthode semi-lagrangienne car comme toute méthode particulaire, les méthodes PICAP sont fortement bruitées. Les méthodes Direct-Implicit donnent des résultats similaires.

Enfin le dernier cas test est celui de l'expansion d'un plasma dans le vide, dont l'étude a été menée dans [60]. A l'interface plasma/vide, le plasma n'est pas quasi-neutre : les électrons ayant une inertie plus grande, ils créent une charge négative à l'avant de l'interface. Ceci implique la formation d'une onde ionique qui vient corriger ce défaut de quasi-neutralité et entraîne l'expansion du plasma. Dans le cas sous-résolu, les méthodes PICAP donnent ici de meilleurs résultats que les méthodes Direct-Implicit, notamment concernant le positionnement de l'interface, tandis que la méthode classique est instable. Ce cas test raide permet de montrer que les méthodes PICAP préservent l'asymptotique contrairement aux méthodes Direct-Implicit.

3.2 Un modèle de Vicsek à deux populations

Dans le chapitre II, nous considérons une variante du modèle de Vicsek à deux populations. Une des deux populations est constituée de particules à l'arrêt (module de vitesse nul) et l'autre de particules en mouvement (module de vitesse égal à 1). Chaque particule interagit avec les particules ayant la même vitesse (0 ou 1) par des interactions de type Vicsek. L'évolution de la k-ème particule, avec position $X_k \in \mathbb{R}^2$ et orientation

$V_k \in \mathbb{S}^1 = \{v \in \mathbb{R}^2, |v| = 1\}$, est donnée par :

$$\begin{aligned}
&\frac{dX_k}{dt} = \eta_k V_k,\\
&dV_k = (\mathrm{Id} - V_k \otimes V_k)(\nu_{\eta_k} \bar{V}_k dt + \sqrt{2d_{\eta_k}} dB_t),\\
&\bar{V}_k = \frac{J_k}{|J_k|}, \quad J_k = \sum_{j,\ \eta_j = \eta_k, |X_j - X_k| \leqslant R} V_j,
\end{aligned}$$

où $\eta_k \in \{0, 1\}$ représente la vitesse de la particule, ν_0 et d_0 sont les paramètres du modèle de Vicsek pour les particules à l'arrêt et ν_1 et d_1 les paramètres du modèle de Vicsek pour les particules en mouvement. De plus, la particule peut passer de l'arrêt au mouvement et du mouvement à l'arrêt : η_k est un processus de saut markovien sur l'espace $\{0, 1\}$. Les taux de transition d'un état à l'autre sont donnés par :

$$g^k_{\eta_k} = \frac{1}{\tau_{\eta_k}} \left(1 + \alpha \frac{1}{N} \sum_{j,\ \eta_j \neq \eta_k,\ |X_j - X_k| \leqslant R} \frac{(1 + V_k \cdot V_j)}{2} \right).$$

Le taux de transition g^k_0 (resp. g^k_1) est celui pour passer de l'arrêt au mouvement (resp. du mouvement à l'arrêt). Ces taux de transitions sont constitués d'une partie constante $\tau^{-1}_{\eta_k}$ et d'une partie d'ordre $\alpha\tau^{-1}_{\eta_k}$ dépendant de l'alignement local avec la population cible : le taux de transition de l'état d'arrêt à l'état de mouvement est accrû par la présence de particules en mouvement ayant une direction voisine. L'objectif est de trouver la dynamique à grande échelle de ce modèle particulaire.

Pour cela, nous introduisons $f_0(x, v, t)$ et $f_1(x, v, t)$, les fonctions de distribution des particules à l'arrêt et en mouvement, dans l'espace des phases $x \in \mathbb{R}^2$ et $v \in \mathbb{S}^1$. Le modèle cinétique champ-moyen associé à la dynamique particulaire est le suivant :

$$\partial_t f_0 = \frac{1}{\delta_V} \mathcal{Q}_0(f_0) + \frac{1}{\delta_E} \mathcal{E}(f_0, f_1), \tag{23}$$

$$\partial_t f_1 + v \cdot \nabla_x f_1 = \frac{1}{\delta_V} \mathcal{Q}_1(f_1) - \frac{1}{\delta_E} \mathcal{E}(f_0, f_1), \tag{24}$$

où $\mathcal{Q}_0$ et $\mathcal{Q}_1$ sont les opérateurs de Vicsek (10)-(11)-(12) avec pour paramètres respectifs ν_0, d_0 et ν_1, d_1 et où $\mathcal{E}(f_0, f_1)$ désigne l'opérateur d'échanges entre les deux populations, dus au passage des particules d'un état à l'autre. Les termes δ_V^{-1} et δ_E^{-1} désignent respectivement l'ordre de grandeur des paramètres des interactions de type Vicsek et l'ordre de grandeur des paramètres d'échanges entre les populations. Nous étudions successivement les deux comportement asymptotiques :

1. $\delta_V \ll \delta_E \ll 1$, lorsque les interactions de type Vicsek au sein d'une population sont prédominantes sur les échanges ;
2. $\delta_E \ll \delta_V \ll 1$, lorsque les échanges sont prédominants sur les interactions de type Vicsek.

Pour étudier la première asymptotique $\delta_V \ll \delta_E \ll 1$, nous fixons tout d'abord $\delta_E = O(1)$ et nous prenons la limite $\delta_V \to 0$. A l'aide de la méthode des moments et de la notion des invariants collisionnels généralisés introduite dans [44], nous obtenons

le système suivant :

$$\begin{aligned}
&\partial_t \rho_0 = \frac{1}{\delta_E} R,\\
&\partial_t \rho_1 + \nabla_x \cdot (c_1 \rho_1 \Omega_1) = -\frac{1}{\delta_E} R,\\
&\rho_0 \partial_t \Omega_0 = \lambda_0 (\mathrm{Id} - \Omega_0 \otimes \Omega_0) \left[-\nabla_x \rho_0 + \frac{\beta_0^{-1}}{\delta_E} S_0 \right],\\
&\rho_1 \partial_t \Omega_1 + \gamma_1 \rho_1 (\Omega_1 \cdot \nabla_x) \Omega_1 = \lambda_1 (\mathrm{Id} - \Omega_1 \otimes \Omega_1) \left[-\nabla_x \rho_1 + \frac{\beta_1^{-1}}{\delta_E} S_1 \right],
\end{aligned}$$

où $\lambda_0 = d_0/\nu_0$, $\lambda_1 = d_1/\nu_1$ et où ρ_0, ρ_1 sont les densités des deux populations, Ω_0, Ω_1 les orientations. Les constantes c_1, γ_1, β_0, β_1 dépendent uniquement des paramètres λ_0 et λ_1. Les termes R, S_0, S_1 sont les termes macroscopiques venant des échanges : ils dépendent donc des densités et des orientations des deux populations. L'objectif est maintenant de trouver la dynamique lorsque $\delta_E \to 0$ dans le système précédent. Par un développement de Hilbert par rapport à δ_E, nous obtenons à l'ordre principal les relations d'équilibre suivantes :

$$R = 0, \quad (\mathrm{Id} - \Omega_0 \otimes \Omega_0) S_0 = 0, \quad (\mathrm{Id} - \Omega_1 \otimes \Omega_1) S_1 = 0.$$

La première nous donne une relation non-linéaire entre les densités ρ_0 et ρ_1 tandis que les deux autres impliquent que les deux populations ont la même direction ou la direction opposée : $\Omega_0 = \Omega_1$ ou $\Omega_0 = -\Omega_1$. L'évolution de ces quantités à l'équilibre est donnée par le système obtenu à l'ordre suivant du développement : ce système fait apparaître les opérateurs d'échanges linéarisés autour des équilibres. Les noyaux de ces opérateurs linéaires sont non réduits à $\{0\}$. En exprimant que le système doit appartenir à l'image de ces opérateurs linéaires, nous obtenons dans le cas $\Omega_0 = \Omega_1$ le système fermé suivant :

$$\begin{aligned}
&\partial_t \rho + \nabla_x \cdot (c_1 k^{-1}(\rho) \Omega) = 0,\\
&(A_1 \rho + (A_0 - A_1) k^{-1}(\rho))\ \partial_t \Omega + A_0 \gamma_1 k^{-1}(\rho)\ (\Omega \cdot \nabla_x) \Omega =\\
&\qquad -\lambda_0 A_1 + (\lambda_1 A_0 - \lambda_0 A_1)(k^{-1})'(\rho))(\mathrm{Id} - \Omega \otimes \Omega)\ \nabla_x \rho,
\end{aligned}$$

où $\rho = \rho_0 + \rho_1$ désigne la densité totale, $\Omega = \Omega_0 = \Omega_1$ la direction commune et k est une fonction croissante de telle sorte que $\rho = k(\rho_1)$.

L'asymptotique $\delta_E \ll \delta_V \ll 1$ est pour le moment moins documentée. Nous montrons toutefois dans le cas où les échanges ne dépendent pas de l'alignement local ($\alpha = 0$) que la relation $\lambda_0 = \lambda_1$ doit être satisfaite pour qu'un équilibre soit atteint. Nous obtenons alors un modèle de Vicsek aux paramètres moyennés différent du modèle obtenu dans l'asymptotique précédente.

3.3 Simulations numériques du système de Vicsek macroscopique

Le chapitre III a pour objet la simulation numérique du modèle macroscopique de Vicsek (P_V^0) et sa concordance avec la dynamique particulaire (7)-(8). Du fait de la non-conservativité du modèle, aucune notion de solution ne peut être privilégiée. C'est pourquoi nous développons plusieurs méthodes numériques, qui constituent autant de définitions, a priori distinctes, de solutions. Nous confrontons numériquement ces méthodes aux simulations du système particulaire de Vicsek sur des cas tests unidimensionnels.

Le modèle de Vicsek unidimensionnel s'écrit ainsi :

$$\partial_t \begin{pmatrix} \rho \\ \theta \end{pmatrix} + A(\rho,\theta)\,\partial_x \begin{pmatrix} \rho \\ \theta \end{pmatrix} = 0, \tag{25}$$

avec

$$A(\rho,\theta) = \begin{bmatrix} \cos\theta & -\rho\sin\theta \\ -\frac{\lambda\sin\theta}{\rho} & c\cos\theta \end{bmatrix}, \tag{26}$$

où ρ est la densité et θ désigne la coordonnée polaire décrivant $\Omega = (\cos\theta, \sin\theta)^T$ et λ et c sont des coefficients du modèle. La première méthode numérique est issue d'une reformulation conservative du système (25)-(26). Cette formulation s'écrit :

$$\partial_t \begin{pmatrix} \rho \\ f_1(\theta) \end{pmatrix} + \partial_x \begin{pmatrix} \rho\cos\theta \\ cf_2(\theta) - \lambda\log(\rho) \end{pmatrix} = 0,$$

avec

$$f_1(\theta) = \log\left|\tan\frac{\theta}{2}\right|, \quad f_2(\theta) = \log|\sin\theta|\,.$$

Cette formulation n'est pas équivalente à (25)-(26) : $f_1(\theta)$ étant une fonction paire, cette formulation ne donne en fait que la valeur absolue de l'angle θ. Une méthode de Roe est alors utilisée pour simuler ce système conservatif et doit être complétée par une autre méthode pour en déterminer le signe de θ.

La deuxième méthode est basée sur la limite de relaxation suivante :

$$\begin{aligned} &\partial_t\rho + \nabla_x\cdot(\rho\Omega) = 0, \\ &\partial_t(\rho\Omega) + c\nabla_x\cdot(\rho\Omega\otimes\Omega) + \lambda\,\nabla_x\rho = \frac{\rho}{\varepsilon}(1-|\Omega|^2)\Omega, \\ &\Omega\in\mathbb{R}^2, \end{aligned}$$

où ε est un paramètre numérique. Nous montrons que la limite formelle de ce système lorsque $\varepsilon \to 0$ est bien le modèle macroscopique de Vicsek. Une méthode de splitting est alors effectuée : à chaque pas de temps, le problème conservatif sans contrainte géométrique,

$$\begin{aligned} &\partial_t\rho + \nabla_x\cdot(\rho\Omega) = 0, \\ &\partial_t(\rho\Omega) + c\nabla_x\cdot(\rho\Omega\otimes\Omega) + \lambda\,\nabla_x\rho = 0, \\ &\Omega\in\mathbb{R}^2, \end{aligned}$$

est résolu par une méthode de Roe puis la vitesse Ω est renormalisée.

Les troisième et quatrième schémas numériques sont tous deux des schémas basés sur la formulation non-conservative initiale (25)-(26). Le troisième est un schéma de type upwind. C'est ce schéma qui sert à définir le signe de θ dans la première méthode. Le quatrième schéma est une variante semi-conservative de la méthode upwind : la discrétisation de l'équation sur l'angle est toujours non-conservative mais la discrétisation de l'équation de la densité est rendue conservative.

La comparaison de ces schémas pour des problèmes de Riemann montre qu'ils coïncident pour des solutions régulières en onde de raréfaction mais qu'ils peuvent ou non coïncider pour des solutions en onde de choc.

La première étape pour comparer la dynamique du système particulaire avec le modèle macroscopique de Vicsek est d'étudier numériquement la convergence de la dynamique particulaire vers l'équilibre donné par la fonction de Von Mises, sur lequel est basé le modèle macroscopique :

$$M_\Omega(v) = \exp\left(\frac{v\cdot\Omega}{\lambda}\right).$$

Pour cela, nous avons comparé la vitesse moyenne théorique à la vitesse moyenne numérique : celles-ci concordent lorsque la densité de particules est suffisamment élevée. Ceci permet de définir le domaine de validité éventuel du modèle macroscopique.

Nous comparons ensuite la dynamique particulaire aux solutions issues des différents schémas numériques associés au modèle macroscopique. Notons que les résultats des simulations particulaires sont moyennés sur plusieurs simulations afin de réduire le bruit numérique. Les cas tests choisis sont des problèmes de Riemann unidimensionnels. Le premier cas test est une onde de choc où les différents schémas macroscopiques ne coïncident pas. Il est observé que la méthode de splitting concorde bien avec les simulations particulaires et que parmi les méthodes testées, elle fournit les meilleurs résultats : la méthode conservative est moins précise que la méthode de splitting tandis que les méthodes upwind et semi-conservative ne sont pas en accord avec le résultat des simulations particulaires. Le second cas test d'onde de contact permet finalement de valider la méthode de splitting. En effet, elle donne dans ce cas un résultat bien différent de l'onde de contact produite par la méthode conservative : c'est une onde composite et c'est cette solution que décrit la méthode particulaire.

Cette étude permet de valider le modèle macroscopique de Vicsek pour les régimes denses.

3.4 La congestion dans un modèle macroscopique pour animaux grégaires

Au chapitre IV, nous étudions semi-formellement, semi-analytiquement un modèle bidimensionnel avec contrainte géométrique et contrainte de densité maximale. Ce modèle s'écrit :

$$\partial_t\rho + \nabla_x\cdot(\rho\Omega) = 0, \tag{27}$$
$$\partial_t\Omega + (\Omega\cdot\nabla_x)\Omega + (\mathrm{Id}-\Omega\otimes\Omega)\nabla_x p(\rho) = 0, \tag{28}$$

où $\Omega\in\mathbb{S}^1$ et $p(\rho)$ est une fonction croissante telle que pour $\rho\ll 1$, $p(\rho)\sim\rho^\gamma$ et telle que $p(\rho)\to+\infty$ quand ρ tend vers la densité maximale ρ^*. La fonction p considérée est :

$$p(\rho) = \frac{1}{\left(\frac{1}{\rho}-\frac{1}{\rho^*}\right)^\gamma}.$$

Ce modèle est dérivé en appendice d'un modèle particulaire où les particules interagissent via une force d'attraction à longue distance modélisant l'agrégation et une force de répulsion à courte distance modélisant les contraintes d'encombrement. La contrainte de densité maximale est contenue dans cette force de répulsion. Des simulations numériques de ce système particulaire montrent la formation d'agrégat de densité élevée suivant la

valeur des paramètres. Par l'intermédiaire d'un modèle cinétique et d'un changement d'échelle hydrodynamique, nous dérivons le modèle (27)-(28) où la force de répulsion est devenue locale. La force d'attraction étant non locale même à grande échelle et donc régulière, elle a été pour le moment écartée afin de se concentrer sur la singularité de la force $\nabla p(\rho)$.

Pour étudier l'impact de cette contrainte de densité maximale, nous introduisons un paramètre ε qui en mesure la raideur : p est remplacé par εp. Ainsi, $\varepsilon p(\rho^\varepsilon)$ tend vers 0 quand $\varepsilon \to 0$ sauf si ρ^ε tend vers la densité maximale ρ^*. Dans ce cas, nous supposons que $\varepsilon p(\rho^\varepsilon)$ converge vers une valeur finie $\bar{p} < +\infty$. Ainsi dans les domaines non-congestionnés $\rho < \rho^*$, le modèle devient un modèle de gaz sans pression avec contrainte géométrique :

$$\begin{aligned} &\partial_t \rho + \nabla_x \cdot \rho\Omega = 0, \\ &\partial_t \Omega + \Omega \cdot \nabla_x \Omega = 0, \\ &|\Omega| = 1. \end{aligned}$$

Dans les domaines congestionnés $\rho = \rho^*$, le modèle devient le système d'Euler incompressible avec contrainte géométrique :

$$\begin{aligned} &\nabla_x \cdot \Omega = 0, \\ &\partial_t \Omega + \Omega \cdot \nabla_x \Omega + (\mathrm{Id} - \Omega \otimes \Omega)\nabla_x \bar{p} = 0, \\ &|\Omega| = 1, \\ &\rho = \rho^*, \end{aligned} \tag{29}$$

où la pression $\bar{p}$ devient le multiplicateur de Lagrange de la contrainte d'incompressibilité (29) et vérifie au sein de ces domaines une équation elliptique. Les solutions régulières en deux dimensions de ce système incompressible sont des mouvements de translation. Ce modèle limite n'est pas complet car la dynamique des interfaces entre la phase congestionnée et la phase non congestionnée n'est pas fournie. Des conditions aux interfaces des domaines sont notamment nécessaires pour déterminer la pression au sein des zones congestionnées.

Pour obtenir des informations sur la dynamique de l'interface, nous avons étudié la limite $\varepsilon \to 0$ des solutions du problème de Riemann unidimensionnel. Pour cela, nous avons considéré une formulation conservative du problème :

$$\partial_t \rho + \partial_x(\rho \cos\theta) = 0, \tag{30}$$

$$\partial_t \Psi(\cos(\theta)) + \partial_x(\Phi(\cos\theta) + \varepsilon p(\rho)) = 0, \tag{31}$$

où $\Psi(\cos\theta) = -\ln|\tan(\theta/2)|$ et $\Phi(\cos\theta) = -\ln|\sin\theta|$. Nous montrons aussi qu'il existe une infinité de formulations conservatives. En l'absence de théorie physique pour sélectionner l'une d'entre elles, nous avons considéré la formulation la plus "simple". De plus, la dynamique des interfaces obtenue par cette méthode dépend de ce choix mais seulement de manière quantitative : la nature des interfaces en est indépendante.

Le système (30)-(31) est strictement hyperbolique. Nous montrons que les champs caractéristiques sont à la limite $\varepsilon \to 0$ vraiment non-linéaires. Par conséquent, les solutions du problème de Riemann associé sont composées d'ondes de raréfaction et d'ondes de choc. Par des considérations géométriques sur les courbes de raréfaction et de choc et sur

leurs intersections, nous obtenons la limite des solutions du problème de Riemann. Elles montrent l'apparition de domaines de densité maximale. De plus, les ondes de raréfaction deviennent à la limite $\varepsilon \to 0$ des ondes de contact et/ou des ondes de "déclusterisation", qui consistent en l'annulation instantanée de la pression. Dans tous les cas, l'analyse du problème de Riemann fournit une solution calculable.

Cette étude permet d'émettre des hypothèses sur la dynamique des différentes interfaces possibles entre les domaines congestionnés (C) où la densité est maximale $\rho = \rho^*$, les domaines non congestionnées (UC) où $0 < \rho < \rho^*$ et les zones vides (V) où $\rho = 0$:

- les interfaces (C)/(V) ou (UC)/(V) sont des ondes de contact avec une pression nulle au bord ;
- les interfaces (C)/(UC) sont des ondes de choc avec un saut de pression, dont la valeur dépend de la formulation conservative ;
- les interfaces (UC)/(UC) sont des ondes de contact.

Ceci permet d'établir la dynamique du modèle limite lorsqu'un seul domaine congestionné est présent. Pour ce qui est de la collision entre deux domaines congestionnés, la limite du problème de Riemann ne permet pas de déterminer l'évolution de l'interface. En effet, lorsque les deux états du problème de Riemann sont de densité maximale, la pression devient instantanément infinie. Le problème est donc mal posé. Pour combler ce manque, une solution unidimensionnelle pour la collision de deux domaines congestionnés de tailles finies est proposée en utilisant la notion de bouchon collant introduite dans [20]. La pression contient alors un delta de Dirac en temps à l'instant de la collision. Toutefois, la géométrie bidimensionnelle de la collision est manquante.

3.5 Simulations numériques du système d'Euler avec congestion

Dans le chapitre V, nous développons des méthodes numériques préservant l'asymptotique de congestion pour le système d'Euler avec contrainte de densité maximale :

$$\partial_t \rho^\varepsilon + \nabla_x \cdot q^\varepsilon = 0, \tag{32}$$

$$\partial_t q^\varepsilon + \nabla_x \cdot \left(\frac{q^\varepsilon \otimes q^\varepsilon}{\rho^\varepsilon} \right) + \nabla_x \varepsilon p(\rho^\varepsilon) = 0, \tag{33}$$

où la loi de pression $p(\rho)$ est définie par :

$$p(\rho) = \frac{1}{\left(\frac{1}{\rho} - \frac{1}{\rho^*}\right)^\gamma}. \tag{34}$$

Lorsque le paramètre ε tend vers 0, le modèle asymptotique comporte là aussi deux phases : une phase de densité maximale $\rho = \rho^*$ où le modèle devient le système d'Euler incompressible et une phase de densité plus faible où $\rho < \rho^*$ où le modèle devient le système des gaz sans pression.

Des solutions unidimensionnelles sont établies pour le problème asymptotique. Comme dans le chapitre IV, nous analysons la limite des solutions du problème de Riemann lorsque $\varepsilon \to 0$: il y alors apparition du vide et de domaines de densité maximale. Pour ce problème, la paramétrisation des courbes de raréfaction et des courbes de choc permet de donner explicitement les solutions des problèmes limites. Les résultats ont déjà été en partie obtenus dans [20] en cherchant directement les solutions du problème de Riemann

pour le problème limite. La discontinuité entre deux états infinis de densité maximale est là encore un problème mal posé. La solution pour la collision de deux domaines congestionnés de taille finies a été donnée dans [20].

Nous ne pouvons simuler directement le problème asymptotique car sa formulation est incomplète. Il manque notamment des conditions aux interfaces pour la collision de deux clusters. Nous développons donc des méthodes numériques pour le problème $\varepsilon > 0$ permettant de capturer la dynamique asymptotique et notamment la dynamique des interfaces entre domaines incompressibles et domaines sans pression. Nous adaptons pour cela deux méthodes proposées pour la limite bas-Mach, la méthode directe [45] et la méthode de jauge [43], et nous les comparons sur plusieurs cas test unidimensionnels.

La méthode directe s'appuie sur une discrétisation temporelle semi-implicite ainsi que sur une décomposition de la pression en une partie explicite $p_0(\rho)$ et une partie implicite $p_1(\rho)$ inversible. La pression implicite $p_1^{n+1} = p_1((n+1)\Delta t)$ au temps $n+1$ est déterminée par l'équation elliptique suivante :

$$\frac{p_1^{-1}(p_1^{n+1}) - \rho^n}{\Delta t} + \nabla_x \cdot q^n - \Delta t\ \varepsilon \Delta_x p_1^{n+1} - \Delta t \Delta_x \left(\frac{q^n \otimes q^n}{\rho^n} + \varepsilon p_0(\rho^n) \right) = 0, \tag{35}$$

obtenue en injectant la discrétisation de l'équation (33) dans la discrétisation de l'équation (32). Une fois la pression p_1^{n+1} calculée, la densité $\rho^{n+1} = \rho((n+1)\Delta t)$ est calculée en inversant la fonction $p_1(\rho)$ puis la quantité de mouvement $q^{n+1} = q((n+1)\Delta t)$ est obtenue par l'équation :

$$q^{n+1} = q^n - \Delta t \left[\nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} + \varepsilon p_0(\rho^n) \right) + \varepsilon \nabla_x (p_1(\rho^{n+1})) \right]. \tag{36}$$

Les discrétisations spatiales sont celles du schéma de Rusanov (ou Lax-Friedrichs local). Le fait de résoudre l'équation sur la pression (35) permet de garantir que la densité reste inférieure à la densité maximale ρ^*.

Le caractère AP de ce schéma est discuté : dans les domaines de densité maximale, il converge vers la bonne solution à condition qu'il capture correctement les conditions d'interfaces. L'absolue stabilité est obtenue grâce à la semi-implicitation dans (36).

La méthode de jauge est un raffinement de la méthode directe. Elle consiste à utiliser la décomposition de Hodge :

$$q = a - \nabla_x \varphi, \quad \nabla_x \cdot a = 0, \tag{37}$$

où a représente la partie incompressible du champ q et $\nabla_x \varphi$ sa partie irrotationnelle. Le schéma est obtenu en insérant cette décomposition dans la méthode précédente. Il donne lieu à la résolution de trois équations elliptiques. Le partie irrotationnelle $\nabla_x \varphi$ est notamment calculée par l'intermédiaire de l'équation suivante :

$$\Delta_x \varphi^{n+1} = \frac{\rho^{n+1} - \rho^n}{\Delta t}. \tag{38}$$

La discrétisation spatiale de l'opérateur Δ_x dans cette équation donne lieu à deux méthodes distinctes, dénommées Gauge 1 et Gauge 2 :

$$\Delta_1 \varphi^{n+1} = \frac{\varphi_{j+1}^{n+1} - 2\varphi_j^{n+1} + \varphi_{j-1}^{n+1}}{(\Delta x)^2}, \quad \Delta_2 \varphi^{n+1} = \frac{\varphi_{j+2}^{n+1} - 2\varphi_j^{n+1} + \varphi_{j-2}^{n+1}}{4(\Delta x)^2}.$$

Plusieurs cas tests unidimensionnels sont étudiés pour comparer ces deux schémas. Ils n'ont pas permis de montrer de réelles différences entre la méthode directe et la méthode Gauge 2. Par contre, la méthode Gauge 1 et la méthode directe semblent avoir des propriétés complémentaires. Lorsque il y a formation d'un domaine congestionné, la méthode directe crée des petites oscillations aux interfaces de ce domaine alors que la méthode Gauge 1 en est dépourvue. La méthode Gauge 1 capture aussi très bien les collisions instantanées de deux domaines congestionnés. Toutefois, la méthode Gauge 1 est bien plus diffusive que la méthode directe notamment lorsqu'il y a formation de vide. Il nous reste donc à analyser ces deux schémas pour comprendre d'où viennent ces différences.

4 Conclusion

Le corps de la thèse est organisé comme suit :

1. Une première partie est consacrée au développement de méthodes numériques pour traiter la limite quasi-neutre dans les plasmas. Elle est composée du chapitre I, où les deux méthodes particulaires pour le système Vlasov-Poisson sont présentées et où une étude numérique détaillée permet de les valider.
2. Dans la seconde partie, le thème central est la limite hydrodynamique avec contrainte géométrique (sur la vitesse des particules) : le chapitre II présente les modèles macroscopiques obtenus lorsque le modèle de Vicsek est étendu au cas où le module de vitesse des particules peut prendre les valeurs 0 ou 1 et le chapitre III apporte une validation numérique du modèle macroscopique de Vicsek en comparant plusieurs schémas numériques à la simulation de la dynamique particulaire.
3. Enfin, la troisième partie traite de la limite de congestion : dans le chapitre IV, nous dérivons le modèle macroscopique avec contrainte de densité maximale et contrainte géométrique à partir d'un modèle particulaire d'agrégation et nous analysons la dynamique des transitions compressible/incompressible qui résultent de la limite de congestion. Le chapitre V propose des méthodes numériques préservant l'asymptotique de congestion pour un système d'Euler avec contrainte de densité maximale.

Plusieurs directions pourront être empruntées pour développer ces travaux. La première consisterait à raffiner des modèles particulaires PIC pour le système Vlasov-Poisson pour réduire le bruit numérique et gagner en précision. Il faudrait aussi compléter l'étude du modèle de Vicsek à deux populations en étudiant la stabilité des équilibres puis comparer numériquement le modèle obtenu avec la dynamique particulaire sous-jacente. Des études devront être aussi menées pour établir théoriquement et numériquement un modèle valide pour les régions peu denses du modèle de Vicsek. Une étude de paramètres du modèle particulaire avec force de répulsion singulière modélisant les contraintes d'encombrement permettrait de déterminer les conditions pour avoir formation de domaines congestionnés. Enfin, une compréhension plus précise des phénomènes d'oscillations dans les méthodes numériques proposées pour l'asymptotique de congestion permettrait éventuellement de corriger ce défaut. Nous renvoyons à la conclusion générale pour d'autres axes de recherches possibles.

Terminons cette introduction en revenant sur la notion de “simplification" employée pour introduire les problèmes asymptotiques (page 15) : les études asymptotiques des

différents modèles de mouvement collectif fournissent des descriptions “simplifiées" et apportent donc des arguments pour discuter leurs validités et leurs utilités du point de vue de la modélisation biologique. Mais ce n'est qu'une première étape qui devra être suivie par la confrontation des différents modèles avec des données expérimentales. Dans cette délicate confrontation, les méthodes numériques mises au point pour capturer les dynamiques macroscopiques des différents modèles pourront s'avérer fort utiles.

Bibliographie

[1] M. Aldana and C. Huepe. Phase transitions in self-driven many-particle systems and related non-equilibrium models : A network approach. *J. Stat. Phys.*, 112(1) :135–153, 2003.

[2] I. Aoki. A simulation study on the schooling mechanism in fish. *Bulletin of the Japanese Society of Scientific Fisheries*, 48(8) :1081–1088, 1982.

[3] A.-A. Arsen'ev. Global existence of a weak solution of Vlasov's system of equations. *U.S.S.R. Comput. Math. Math. Phys.*, 15(1) :131– 143, 1975.

[4] A. Aw, A. Klar, T. Materne, and M. Rascle. Derivation of continuum traffic flow models from microscopic follow-the-leader models. *SIAM J. Appl. Math.*, 63 :259–278, 2002.

[5] A. Aw and M. Rascle. Resurrection of “second order" models of traffic flow. *SIAM J. Appl. Math.*, 60 :916–938, 2000.

[6] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, et al. Empirical investigation of starling flocks : a benchmark study in collective animal behaviour. *Animal Behaviour*, 76(1) :201–215, 2008.

[7] R. Belaouar, N. Crouseilles, P. Degond, and E. Sonnendrücker. An asymptotically stable semi-lagrangian scheme in the quasi-neutral limit. *J. Sci. Comp.*, 41(3) :341–365, 2009.

[8] N. Bellomo and C. Dogbe. On the modelling crowd dynamics from scaling to hyperbolic macroscopic models. *Math. Model Meth. Appl. Sci.*, 18 :1317–1345, 2008.

[9] M. Bennoune, M. Lemou, and L. Mieussens. Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier-Stokes asymptotics. *J. Comput. Phys.*, 227(8) :3781–3803, 2008.

[10] M. Bennoune, M. Lemou, and L Mieussens. An asymptotic preserving scheme for the Kac model of the Boltzmann equation in the diffusion limit. *Continuum Mech. Therm.*, 21(5) :401–421, 2009.

[11] F. Berthelin. Existence and weak stability for a pressureless model with unilateral constraint. *Math. Model Meth. Appl. Sci.*, 12(2) :249–272, 2002.

[12] F. Berthelin, P. Degond, M. Delitala, and M. Rascle. A model for the formation and the evolution of traffic jams. *Arch. Rational Mech. Anal.*, 187(2) :185–220, 2008.

[13] F. Berthelin, P. Degond, V. Le Blanc, S. Moutari, M. Rascle, and J. Royer. A traffic-flow model with constraints for the modeling of traffic jams. *Math. Model Meth. Appl. Sci.*, 18 :1269–1298, 2008.

[14] E. Bertin, M. Droz, and G. Grégoire. Boltzmann and hydrodynamic description for self-propelled particles. *Phys. Rev. E.*, 74 :022101, 2006.

[15] A.L. Bertozzi, J.A. Carrillo, and T. Laurent. Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. *Nonlinearity*, 22 :683–710, 2009.

[16] A.L. Bertozzi and T Laurent. Finite-time blow-up of solutions of an aggregation equation in $\mathbb{R}$. *Commun. Math. Phys.*, 274 :717–735, 2007.

[17] C.K. Birdsall and A.B. Langdon. *Plasma physics via computer simulation*. Inst of Physics Pub Inc, 2004.

[18] E. Bonabeau, G. Theraulaz, J.L. Deneubourg, S. Aron, and S. Camazine. Self-organization in social insects. *Trends in Ecology & Evolution*, 12(5) :188–193, 1997.

[19] F. Bouchut. On zero pressure gas dynamics. *Advances in kinetic theory and computing : selected papers, Ser. Adv. Math. Appl. Sci.*, 22 :171–190, 1994.

[20] F. Bouchut, Y. Brenier, J. Cortes, and J.-F. Ripoll. A hierarchy of models for two-phase flows. *J. Nonlinear Sci.*, 10(6) :639–660, 2000.

[21] Y. Brenier. Convergence of the Vlasov-Poisson system to the incompressible Euler equations. *Comm. Part. Diff. Equat.*, 25 :737, 2000.

[22] Y. Brenier and E. Grenier. Sticky particles and scalar conservation laws. *SIAM J. Num. Anal.*, 35(6) :2317–2328, 1998.

[23] J. Buhl, D.J.T. Sumpter, I.D. Couzin, J.J. Hale, E. Despland, E.R. Miller, and S.J. Simpson. From disorder to order in marching locusts. *Science*, 312(5778) :1402–1406, 2006.

[24] V. Calvez and J.A. Carrillo. Volume effects in the Keller–Segel model : energy estimates preventing blow-up. *J. Math. Pure Appl.*, 86(2) :155–175, 2006.

[25] S. Camazine, J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-organization in biological systems*, volume Princeton studies in complexity. Princeton, NJ : Princeton studies in complexity, 2001.

[26] J.A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani. Asymptotic flocking dynamics for the kinetic Cucker-Smale model. *preprint*, 2009.

[27] J.A. Carrillo, T. Goudon, and P. Lafitte. Simulation of fluid and particles flows : Asymptotic preserving schemes for bubbling and flowing regimes. *J. Comput. Phys.*, 227(16) :7929–7951, 2008.

[28] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, and A. Procaccini. The STARFLAG handbook on collective animal behaviour : 2. Three-dimensional analysis. *Animal Behaviour*, 76(1) :237–248, 2008.

[29] A. Cavagna, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. The STARFLAG handbook on collective animal behaviour : 1. Empirical methods. *Animal Behaviour*, 76(1) :217–236, 2008.

[30] F.F. Chen. *Introduction to plasma physics and controlled fusion*. Plenum Press, 1974.

[31] R.M. Colombo and M.D. Rosini. Pedestrian flows and non-classical shocks. *Mathematical methods in the applied sciences*, 28(13) :1553–1568, 2005.

[32] F. Coron and B. Perthame. Numerical passage from kinetic to fluid equations. *SIAM J. Numer. Anal.*, 28(1) :26–42, 1991.

[33] G.H. Cottet and P.A. Raviart. Particle methods for the one-dimensional Vlasov-Poisson equations. *SIAM J. Numer. Anal.*, 21 :52–76, 1984.

[34] I.D. Couzin and J. Krause. Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, 32(1), 2003.

[35] I.D. Couzin, J. Krause, R. James, G.D. Ruxton, and N.R. Franks. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1) :1–11, 2002.

[36] P. Crispel, P. Degond, J.-G. Liu, and M.-H. Vignal. Analysis of an asymptotic preserving scheme for the Euler-Poisson system in the quasineutral limit. *SIAM J. Numer. Anal.*, 46(3) :1298–1322, 2008.

[37] P. Crispel, P. Degond, and M.-H. Vignal. An asymptotic preserving scheme for the two-fluid Euler-Poisson model in the quasineutral limit. *J. Comput. Phys.*, 223(1) :208, 2007.

[38] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5) :852–862, 2007.

[39] F. Cucker and S. Smale. On the mathematics of emergence. *Japan J. Math.*, 2(1) :197–227, 2007.

[40] P. Degond. Macroscopic limits of the Boltzmann equation : a review. In P. Degond, L. Pareschi, and R. Russo, editors, *Modeling and computational methods for kinetic equations*, volume Modeling and Simulation in Science of *Engineering and Technology Series*, pages 3–57. Birkhaüser, Boston, 2003.

[41] P. Degond, S. Hirstoaga, and M.-H. Vignal. The Vlasov model under large magnetic fields in the low-Mach number regime. *preprint, arXiv :0905.2400*.

[42] P. Degond, S. Jin, and M. Tang. On the Time Splitting Spectral Method for the Complex Ginzburg-Landau Equation in the Large Time and Space Scale Limit. *SIAM J. Sci. Comput.*, 30(5) :2466–2487, 2008.

[43] P. Degond, S. Jin, and J. Yuming. Mach-number uniform asymptotic-preserving gauge schemes for compressible flows. *Bulletin of the Institute of Mathemtics, Academia Sinica*, 2(4) :851, 2007.

[44] P. Degond and S. Motsch. Continuum limit of self-driven particles with orientation interaction. *Math. Model. Meth. Appl. Sci.*, 18 :1193–1215, 2008.

[45] P. Degond and M. Tang. All speed scheme for the low mach number limit of the Isentropic Euler equation. *to appear in Math. Model. Meth. Appl. Sci.*, 2010.

[46] P. Degond and T. Yang. Diffusion in a continuum model of self-propelled particles with alignment interaction. *preprint*, 2010.

[47] J.L. Deneubourg and S. Goss. Collective patterns and decision-making. *Ethology, ecology and Evolution*, 1(4) :295–311, 1989.

[48] M.R. D'Orsogna, Y.L. Chuang, A.L. Bertozzi, and L.S. Chayes. Self-propelled particles with soft-core interactions : patterns, stability, and collapse. *Phys. Rev. Lett.*, 96(10) :104302, 2006.

[49] U. Erdmann, W. Ebeling, and A.S. Mikhailov. Noise-induced transition from translational to rotational motion of swarms. *Phys. Rev. E*, 71(5) :51904, 2005.

[50] F. Filbet, P. Laurencot, and B. Perthame. Derivation of hyperbolic models for chemosensitive movement. *J. Math. Biol.*, 50(2) :189–207, 2005.

[51] F. Filbet, E. Sonnendrucker, and P. Bertrand. Conservative numerical schemes for the Vlasov equation. *J. Comput. Phys.*, 172(1) :166–187, 2001.

[52] A. Frouvelle. A continuous model for alignment of self-propelled particles with anisotropy and density-dependent parameters. *preprint*, 2010.

[53] E. Gabetta, L. Pareschi, and G. Toscani. Relaxation schemes for nonlinear kinetic equations. *SIAM J. Numer. Anal.*, 34(6) :2168–2194, 1997.

[54] K. Ganguly and H.D. Victory Jr. On the convergence of particle methods for multidimensional Vlasov-Poisson systems. *SIAM J. Numer. Anal.*, 26(2) :249–288, 1989.

[55] J. Gautrais, C. Jost, M. Soria, A. Campo, S. Motsch, R. Fournier, S. Blanco, and G. Theraulaz. Analyzing fish movement as a persistent turning walker. *J. Math. Biol.*, 58(3) :429–445, 2009.

[56] F. Golse, S. Jin, and C.D. Levermore. The convergence of numerical transfer schemes in diffusive regimes I : discrete-ordinate method. *SIAM J. Numer. Anal.*, 36(5) :1333–1369, 1999.

[57] G. Grégoire, H. Chaté, and Y. Tu. Moving and staying together without a leader. *Phys. Nonlinear Phenom.*, 181 :157–170, 2003.

[58] E. Grenier. Defect measures of the Vlasov-Poisson system in the quasineutral regime. *Comm. Part. Differ. Equat.*, 20(7) :1189–1216, 1995.

[59] E. Grenier. Global existence for zero pressure gas dynamics.(Existence globale pour le système des gaz sans pression.). *C.R. Acad. Sci., Paris, Sér. I*, 321(2) :171–174, 1995.

[60] T. Grismayer, P. Mora, J.C. Adam, and A. Héron. Electron kinetic effects in plasma expansion and ion acceleration. *Phys. Rev. E*, 77 :066407, 2008.

[61] G. Grégoire and H. Chaté. Onset of collective and cohesive motion. *Phys. Rev. Lett.*, 92 :025702, 2004.

[62] S.Y. Ha and J.G. Liu. A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. *Comm. Math. Sci.*, 7(2) :297–325, 2009.

[63] S.Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic descriptions of flocking. *Kinet. Relat. Models*, 1(3) :415–435, 2008.

[64] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of modern physics*, 73(4) :1067–1141, 2001.

[65] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51 :4282–4286, 1985.

[66] R.W. Hockney and J.W. Eastwood. *Computer simulation using particles*. Taylor & Francis, 1988.

[67] A. Huth and C. Wissel. The simulation of the movement of fish schools*. *J. Theor. Biol.*, 156(3) :365–385, 1992.

[68] M. Ishii. *Thermo-Fluid Dynamic :Theory of Two-Phase Flow*. Eyrolles, Paris, 1975.

[69] S. Jin. Efficient asymptotic-Preserving (AP) schemes for some multiscale kinetic equations. *SIAM J. Sci. Comp.*, 21 :441, 1999.

[70] S. Jin and L. Pareschi. Asymptotic-Preserving (AP) Schemes for Multi-scale Kinetic Equations : A Unified Approach. In Basel Birkhäuser, editor, *Hyperbolic problems : theory, numerics, applications, Vol. I, II (Magdeburg, 2000)*, volume 141 of *Internat. Ser. Numer. Math.*, page 573–582, 2001.

[71] S. Jin, L. Pareschi, and G. Toscani. Uniformly accurate diffusive relaxation schemes for multiscale transport equations. *SIAM J. Numer. Anal.*, 38(3) :913–936, 2001.

[72] S. Jin, P. Pareschi, and G. Toscani. Diffusive relaxation schemes for multiscale, discrete-velocity kinetic equations. *SIAM J. Numer. Anal.*, 35(6) :2405–2439, 1998.

[73] S. Jin and Z. Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Comm. Pure Appl. Math*, 48(3) :235–276, 1995.

[74] S. Klainerman and A. Majda. Compressible and incompressible fluids. *Communications in Pure Applied Mathematics*, 35 :629–651, 1982.

[75] A. Klar. An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit. *SIAM J. Numer. Anal.*, 35(3) :1073–1094, 1998.

[76] A. Klar. An asymptotic preserving numerical scheme for kinetic equations in the low Mach number limit. *SIAM J. Numer. Anal.*, 36(5) :1507–1527, 1999.

[77] A. Klar and A. Unterreiter. Uniform stability of a finite difference scheme for transport equations in diffusive regimes. *SIAM J. Numer. Anal.*, 40(3) :891–913, 2003.

[78] R. Klein. Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics I : one-dimensional flow. *J. Comput. Phys.*, 121(2) :213–237, 1995.

[79] R. Kowalczyk. Preventing blow-up in a chemotaxis model. *J. Math. Anal. Appl.*, 305(2) :566–588, 2005.

[80] A.B. Langdon, B.I. Cohen, and A. Friedman. Direct implicit large time-step particle simulation of plasmas. *J. Comput. Phys.*, 51(107-138) :47–86, 1983.

[81] P.G. LeFloch. Shock waves for nonlinear hyperbolic systems in nonconservative form. *Institute for Mathematics and its Application, Minneapolis*, Preprint 593, 1989.

[82] M. Lemou and L. Mieussens. A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. *SIAM J. Sci. Comput*, 31(1) :334–368, 2008.

[83] H. Levine, W.J. Rappel, and I. Cohen. Self-organization in systems of self-propelled particles. *Phys. Rev. E*, 63(1) :17101, 2000.

[84] M.J. Lighthill and G.B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 317–345, 1955.

[85] P.L. Lions and N. Masmoudi. On a free boundary barotropic model. In *Ann. Inst. H. Poincaré Anal. Non Linaire*, volume 16, pages 373–410, 1999.

[86] PL Lions and B. Perthame. Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Invent. Math.*, 105(1) :415–430, 1991.

[87] J.G. Liu and L. Mieussens. Analysis of an asymptotic preserving scheme for linear kinetic equations in the diffusion limit. *preprint*, 2009.

[88] N. Masmoudi. From Vlasov-Poisson system to the incompressible Euler system. *Comm. Pure Appl. Math.*, 26(9) :1913–1928, 2001.

[89] R.J. Mason. Implicit moment particle simulation of plasmas. *J. Comp. Phys.*, 41(2) :233–244, 1981.

[90] B. Maury, A. Roudneff-Chupin, and F. Santambrogio. A macroscopic crowd motion model of gradient flow gradient. *preprint*.

[91] B. Maury and J. Venel. A mathematical framework for a crowd motion model. *Comptes rendus-Mathématique*, 346(23-24) :1245–1250, 2008.

[92] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *J. Math. Biol.*, 38(6) :534–570, 1999.

[93] A. Mogilner, L. Edelstein-Keshet, L. Bent, and A. Spiros. Mutual interactions, potentials, and individual distance in a social aggregation. *J. Math. Biol.*, 47(4) :353–389, 2003.

[94] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz. Experimental study of the behavioural mechanisms of spatial self-organization in pedestrian crowds. *Proc. Royal Society B*, 276 :2755, 2009.

[95] M. Moussaïd, D. Helbing, and G. Theraulaz. Simple heuristics that make human crowds smart. in prep., 2010.

[96] L. Pareschi and G. Russo. Time relaxed Monte Carlo methods for the Boltzmann equation. *SIAM J. Sci. Comput.*, 23(4) :1253–1273, 2002.

[97] J.K. Parrish and L. Edelstein-Keshet. Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284(5411) :99, 1999.

[98] J.K. Parrish, S.V. Viscido, and D. Grunbaum. Self-organized fish schools : an examination of emergent properties. *The biological bulletin*, 202(3) :296, 2002.

[99] C. Parés. Numerical methods for nonconservative hyperbolic systems : a theoretical framework. *SIAM J. Numer. Anal.*, 44 :300–321, 2006.

[100] F. Peruani, A. Deutsch, and M. Bär. A mean-field theory for self-propelled particles interacting by velocity alignment mechanisms. *The European Physical Journal-Special Topics*, 157(1) :111–122, 2008.

[101] K. Pfaffelmoser. Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. *J. Differ. Equat.*, 95(2) :281–303, 1992.

[102] M.H. Pillot, J. Gautrais, P. Arrufat, I. D. Couzin, R. Bon, and J.L. Deneubourg. Merinos' sheep wait for the others to commit. in prep., 2010.

[103] M.H. Pillot, J. Gautrais, J. Gouello, P. Michelena, A. Sibbald, and R. Bon. Moving together : incidental leaders and naïve followers. *Behav. Proc.*, 83(3) :235–241, 2010.

[104] C.W. Reynolds. Flocks, herds and schools : A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34. ACM New York, NY, USA, 1987.

[105] J. Shen. Cucker–smale flocking under hierarchical leadership. *SIAM J. Appl. Math.*, 68(3) :694–719, 2007.

[106] A.-S. Snitzman. Topics in propagation of chaos, Cours de l'Ecole d'été de Saint-Flour, Lecture notes in mathematics vol. 1464, Springer, 1989.

[107] E. Sonnendrücker, J. Roche, P. Bertrand, and A. Ghizzo. The semi-Lagrangian method for the numerical resolution of the Vlasov equation. *J. Comput. Phys.*, 149(2) :201–220, 1999.

[108] H. Spohn. *Large Scale Dynamics of Interacting Particles*. Springer-Verlag New York, 1991.

[109] H.B. Stewart and B. Wendroff. Two-phase flow : models and methods. *J. Comput. Phys*, 56(3) :363–409, 1984.

[110] D.J.T. Sumpter. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 361(1465) :5, 2006.

[111] C.M. Topaz and A.L. Bertozzi. Swarming patterns in a two-dimensional kinematic model for biological groups. *SIAM J. Appl. Math.*, 65(1) :152–174, 2005.

[112] C.M. Topaz, A.L. Bertozzi, and M.A. Lewis. A Nonlocal Continuum Model for Biological Aggregation. *Bull. Math. Biol.*, 68(7) :1601–1623, 2006.

[113] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys. Rev. Lett.*, 75(6) :1226–1229, 1995.

[114] C. Villani. Limite Champ-Moyen, Cours de DEA 2001-2002.

[115] E. Weinan, Y.G. Rykov, and Y.G. Sinai. Generalized variational principles, global weak solutions and behavior with random initial data for systems of conservation laws arising in adhesion particle dynamics. *Comm. Math. Phys.*, 177(2) :349–380, 1996.

[116] Y.B. Zeldovich. Gravitational instability : an approximate theory for large density perturbations. *Astron. Astrophys*, 5(84) :168, 1970.

Première partie : Limite quasi-neutre dans les plasmas

Chapitre I

Méthode Particle-In-Cell préservant l'asymptotique quasi-neutre du système Vlasov-Poisson

Ce chapitre a été écrit en collaboration avec Pierre Degond, Fabrice Deluzet, An-Bang Sun et Marie-Hélène Vignal. Il fait l'objet de l'article suivant :

Pierre Degond, Fabrice Deluzet, Laurent Navoret, An-Bang Sun et Marie-Hélène Vignal, ***Asymptotic-Preserving Particle-In-Cell method for the Vlasov-Poisson system near quasineutrality****, J. Comp. Phys., 229(16) :5630-5652, 2010.*

La suite de ce chapitre est écrite en anglais.

Asymptotic-Preserving Particle-In-Cell method for the Vlasov-Poisson system near quasineutrality

Abstract: This paper deals with the numerical resolution of the Vlasov-Poisson system in a nearly quasineutral regime by Particle-In-Cell (PIC) methods. In this regime, classical PIC methods are subject to stability constraints on the time and space steps related to the small Debye length and large plasma frequency. Here, we propose an "Asymptotic-Preserving" PIC scheme which is not subject to these limitations. Additionally, when the plasma period and Debye length are small compared to the time and space steps, this method provides a consistent PIC discretization of the quasineutral Vlasov equation. We perform several one-dimensional numerical experiments which provide a solid validation of the method and its underlying concepts, and compare the method with classical PIC and Direct-Implicit methods.

1 Introduction

The impact of plasmas and more generally, charged-particle fluids on the human environment is constantly increasing, due to their importance in such domains as industrial processes, energy, lighting, air or water cleaning, etc. Because of the large variety of physical situations and the complex multiscale character of most plasma phenomena,

the numerical simulation of plasmas still represents an important challenge for the scientific community. Roughly speaking, according to the physical context, two large classes of mathematical models can be used: fluid models and kinetic ones. This paper is concerned with collisionless plasmas for which a kinetic description is required.

The basic kinetic model for plasma simulations is the Vlasov equation, coupled with the electromagnetic field equations. The Vlasov equation is posed on a 6-dimensional phase-space (3 space dimensions and 3 velocity dimensions) plus time. For this reason, particle methods have been preferred to grid-based (eulerian) methods, as they allow a coarse, yet sufficiently precise, description of the phase space. In Particle-In-Cell methods, the coupling between the particles and the field is implemented through the introduction of a space grid. Charge and current densities are assigned from the particles to the grid. Then, the fields are computed using finite difference methods on the grid and then, interpolated back to the positions of the particles. We refer to the two celebrated books [3, 27] for an overview of these methods. Recently, grid-based eulerian simulations have received a great deal attention [1, 6, 16–18, 38, 40] but particle methods are still the number-one method used for the numerical simulation of plasma kinetic models. The convergence of PIC methods has been mathematically investigated in [9, 19].

One of the very basic but very important problem in plasma simulations is the handling of quasineutrality. Indeed, the electrostatic force tends to restore the local charge neutrality of the plasma. The Debye length and plasma periods [7, 30] set the typical space and time scales at which this restoring force acts. The Debye length measures the typical scale of charge unbalances in the plasma whereas the electron plasma period characterizes the oscillation period of the particles when a departure to quasineutrality occurs. Usually, both these space and time scales are very short compared to the typical scales of the phenomena under investigation. In such situation, the plasma is locally quasineutral. These very short time scales make numerical simulations very time consuming. Indeed, standard explicit PIC methods require a stability condition which guarantees that the space and time steps are smaller than the Debye length and electron plasma period.

Lots of efforts have been devoted to the search for implicit PIC schemes which would be free of such constraints. There are basically two classes of implicit PIC methods : the direct implicit method [8, 31] and the implicit moment method [33, 34]. In the direct implicit method, an implicit algorithm for the advancement of the particles is introduced. However, since a full implicit resolution of the particle positions and of the fields is virtually impossible, a two-step predictor-corrector approximation is practically implemented. In the implicit moment method, a prediction of the value of the fields at the next time step is done through the use of the moment equations. Numerous extensions of these methods can be found in the literature, especially concerning the coupling with the Maxwell equations [4, 25, 32, 35, 36, 41].

Recently, new classes of methods for singular perturbation problems have emerged. These are the so-called Asymptotic-Preserving (AP) methods. Let (S_λ) be a singularly perturbed system and (S_0) the limit system when $\lambda \to 0$. Here, we give a definition of an AP method.

Definition 1. *An Asymptotic Preserving scheme for (S_λ) in the limit $\lambda \to 0$ is a scheme which is consistent with (S_λ) when the numerical parameters (e.g. Δx, Δt) resolve the scales associated with the small parameter λ and which is consistent with (S_0) when $\lambda \to 0$ with Δt, Δx staying of order one.*

The concept of AP method originates from the work of Shi Jin for multiscale kinetic equations [28]. In our case, (S_λ) is the Vlasov-Poisson system and (S_0) is the quasineutral Vlasov system.

The concept of AP method is particularly interesting when λ is not uniformly small. For instance, at a plasma edge, the parameter λ, which depends on the local value of the plasma density, can vary by several orders of magnitude from $\lambda \ll 1$ to $\lambda = O(1)$. In this case, the original problem (S_λ) must be solved in the region where $\lambda = O(1)$ and the limit problem (S_0), where $\lambda \ll 1$. With classical method, this situation requires a model coupling methodology to connect the two models. However, model coupling methods involve a certain level of arbitrariness, such as the location of the coupling interface or the expression of the coupling terms. Their implementation can also be quite complex with the need to adapt the mesh to the geometry of the interface. The AP method allows the computation of the two regions $\lambda = O(1)$ and $\lambda \ll 1$ with the same and unique method. The AP scheme automatically shifts from the (S_λ) model to the (S_0) model wherever λ becomes small, without any need to reduce the time and space steps. This results in a considerably more robust numerical code. We shall see an example of this situation with the simulation of the expansion of an ion slab at section 4.4.

In this paper, we propose an Asymptotic Preserving PIC method (or PICAP method) for the Vlasov-Poisson equation in the quasineutral limit. Previous works on AP methods for the quasineutral limit have been devoted to the Euler-Poisson problem [10, 12, 14] and to Eulerian schemes for the Vlasov-Poisson problem [2]. The quasineutral limit in plasmas has been theoretically investigated in [5, 11, 15, 20, 21, 24, 37].

The present work is a follow-up of a previous work [13], where a first variant of the method (the PICAP-1 method) was presented and tested on a one-species model of a perturbed Maxwellian plasma. In the present work, we introduce a new, simpler variant of this method (the PICAP-2 method) and considerably broaden the list of test problems. These are,

1. a steady-state simulation to quantify the dissipation of the scheme,
2. one and two species perturbations of a Maxwellian plasma with a particular consideration of the Landau damping test problem,
3. the Bump-on-Tail instability,
4. the expansion of an ion slab, following [22, 23].

We will also compare our method to various variants of the Direct-Implicit method.

The method relies on the remark that the equation allowing the computation of the potential in the quasineutral Vlasov problem is very different from the Poisson equation. Indeed, the former is an elliptic equation found from the divergence free condition on the current, a consequence of the quasineutrality. To build an AP scheme, it is necessary to find a unified framework for both the Poisson equation and the quasineutral potential equation. This is done by reformulating the Poisson equation into a strictly equivalent equation which explicitly contains the quasineutral potential equation as a particular case when $\lambda = 0$. This equation is differential in both time and space, and specifically second order in time. An implicit discretization of this equation is combined with a semi-implicit discretization of the particle trajectories and yields an Asymptotic-Preserving scheme for the Vlasov-Poisson problem in the quasineutral limit. Since the reformulation of the Poisson equation leads to a fully equivalent problem, there is no other approximation

involved in our method than purely numerical ones (i.e. associated to the time and space discretization). To our knowledge, previous implicit methods such as the direct implicit or the moment implicit ones (see references above) have not been analyzed in view of this Asymptotic Preservation property, and we will show that they are not consistant with the quasineutral limit problem when $\lambda \to 0$.

This paper is organized as follows. In section 2, we present the two-fluid Vlasov-Poisson system model and its quasineutral limit and we derive the reformulated Poisson equation. From this reformulation, in section 3, we build up the Asymptotic-Preserving Particle-In-Cell (PICAP) method. Section 4 is devoted to the discussion of the numerical results: comparisons between the classical and asymptotic preserving schemes are provided in a one-dimensional geometry. We consider the various test problems as listed above. The results confirm that the AP strategy remains stable and provides the expected results even if the time and space discretizations do not resolve the Debye length and plasma periods. This shows that the method can powerfully deal with stiff problems when the stiffness results from the quasineutrality constraint. By contrast, the Direct-Implicit method show lower accuracy in stiff problems such as the expansion of an ion slab problem.

2 The Vlasov-Poisson system and its quasineutral limit

In this section, we present the two-fluid Vlasov-Poisson system and its quasineutral limit.

2.1 The Vlasov-Poisson system

We are interested in the kinetic description of a two-fluid plasma constituted of electrons and one ion species. Then, ions and electrons are described by their distribution function respectively denoted by $f_i(x, v, t)$ and $f_e(x, v, t)$, where the position and velocity variables x and v are such that $(x, v) \in \Omega \times \mathbb{R}^d$, with $\Omega \subset \mathbb{R}^d$ and $d = 1$, 2, or 3 and $t \geq 0$ is the time. The two-fluid Vlasov-Poisson system is written:

$$\partial_t f_i + v \cdot \nabla_x f_i - \frac{e}{m_i} \nabla_x \phi \cdot \nabla_v f_i = 0, \tag{I.1}$$

$$\partial_t f_e + v \cdot \nabla_x f_e + \frac{e}{m_e} \nabla_x \phi \cdot \nabla_v f_e = 0, \tag{I.2}$$

where we denote by $e > 0$ the positive elementary charge, by $m_{i,e}$ the ion and electron masses and by ϕ the electric potential. ϕ is given by the Poisson equation:

$$-\Delta_x \phi = \frac{e}{\varepsilon_0}(n_i - n_e), \tag{I.3}$$

where ε_0 is the vacuum permittivity, and $n_{i,e}$ are the ion and electron densities, given by

$$n_i(x,t) = \int_{\mathbb{R}^d} f_i(x, v, t) dv, \qquad n_e(x,t) = \int_{\mathbb{R}^d} f_e(x, v, t) dv.$$

The two important physical scales for this model are the Debye length λ_D and the electron plasma frequency ω_p given by:

$$\lambda_D = \left(\frac{\varepsilon_0 k_B T_0}{e^2 n_0}\right)^{1/2}, \qquad \omega_p = \left(\frac{n_0 e^2}{\varepsilon_0 m_e}\right)^{1/2},$$

where k_B is the Boltzmann constant, n_0 is the plasma density scale ($n_0 \backsim n_i \backsim n_e$), T_0 is the plasma temperature scale. We note that an ion plasma frequency can be defined (replacing m_e by m_i). However, this parameter is smaller than ω_p because of the large ion to electron mass ratio. The electron plasma period is defined by $\tau_p = 1/\omega_p$.

The situation where both the Debye length and electron plasma period are very small compared with the typical macroscopic length and time scales is called the quasineutral regime because the local electric charge is very small (of order $k^2\lambda_D^2$ where k is the local inverse gradient length of the electric potential). Simultaneously, the electron plasma period becomes very small as well, so that when local charge unbalances occur (at the scale of the Debye length), very high plasma frequency oscillations are triggered.

In order to study the quasineutral model, we introduce the following scaling of the Vlasov-Poisson problem. The scaled variables are given by

$$\bar{x} = \frac{x}{x_0},\ \bar{v} = \frac{v}{v_0},\ \bar{t} = \frac{v_0}{x_0}t,\ \bar{n}_{i,e} = \frac{n_{i,e}}{n_0},\ \bar{f}_{i,e} = \frac{v_0}{n_0}f_{i,e}, \bar{\phi} = \frac{e\phi}{k_B T_0}, \tag{I.4}$$

where $x_0 > 0$ is the typical length of the problem, and $v_0 \in \mathbb{R}$ is the thermal ion velocity scale given by $v_0 = (k_B T_0/m_i)^{1/2}$. Inserting this scaling into equations (I.1)-(I.3) and omitting the bars, we get the following scaled two-fluid Vlasov-Poisson model

$$\partial_t f_i + v \cdot \nabla_x f_i - \nabla_x \phi \cdot \nabla_v f_i = 0, \tag{I.5}$$

$$\partial_t f_e + v \cdot \nabla_x f_e + \frac{1}{\varepsilon}\nabla_x \phi \cdot \nabla_v f_e = 0, \tag{I.6}$$

$$-\lambda^2 \Delta_x \phi = (n_i - n_e), \quad n_{i,e} = \int f_{i,e}\, dv. \tag{I.7}$$

where $\lambda = \lambda_D/L$ is the scaled Debye length and $\varepsilon = m_e/m_i$ is the electron to ion mass ratio. Note that the scaled plasma frequency is given by $\omega = \omega_p x_0/v_0 = 1/(\sqrt{\varepsilon}\lambda)$. In the following, we investigate the limit $\lambda \to 0$ which leads to the quasineutral limit.

In the forthcoming test cases, we will also be interested in the one-species case, considering only electrons while ions are supposed static (due to their very large mass). The scaled one-species Vlasov-Poisson problem is written:

$$\partial_t f + v \cdot \nabla_x f + \nabla_x \phi \cdot \nabla_v f = 0, \tag{I.8}$$

$$-\lambda^2 \Delta_x \phi = (n_0 - n), \quad n = \int f\, dv\,, \tag{I.9}$$

where n_0 is supposed to be a uniform ion background density.

2.2 Reformulation of the Poisson equation

The quasineutral limit consists in letting $\lambda \to 0$ in the scaled Vlasov-Poisson system (I.5)-(I.7). Obviously, when $\lambda = 0$, we lose the possibility of using the Poisson

equation (I.7) to compute the potential ϕ since (I.7) reduces to the quasineutrality constraint $n_i = n_e$. For this reason, we introduce a reformulation of the problem which provides a more convenient approach to the quasineutral limit.

To this aim, we take the velocity moments of eqs. (I.5) and (I.6) and obtain the mass and momentum conservation equations:

$$\partial_t n_i + \nabla_x \cdot (nu)_i = 0, \tag{I.10}$$
$$\partial_t (nu)_i + \nabla_x \cdot S_i = -n_i \nabla_x \phi, \tag{I.11}$$
$$\partial_t n_e + \nabla_x \cdot (nu)_e = 0, \tag{I.12}$$
$$\partial_t (nu)_e + \nabla_x \cdot S_e = \frac{n_e}{\varepsilon} \nabla_x \phi, \tag{I.13}$$

where $(nu)_{i,e}$, the ion and electron momenta and $S_{i,e}$, the (specific) momentum fluxes are given by

$$(nu)_{i,e} = \int_{\mathbb{R}^d} f_{i,e}(x,v,t)\, v\, dv, \qquad S_{i,e} = \int_{\mathbb{R}^d} f_{i,e}(x,v,t)\, v \otimes v\, dv,$$

and the symbol $\otimes$ denotes the tensor product. Taking the differences of (I.10) and (I.12) on the one hand, and of (I.11) and (I.13) on the other hand, we get the continuity and current equations respectively:

$$\partial_t (n_i - n_e) + \nabla_x \cdot ((nu)_i - (nu)_e) = 0. \tag{I.14}$$
$$\partial_t ((nu)_i - (nu)_e) + \nabla_x \cdot (S_i - S_e) = -(n_i + \frac{n_e}{\varepsilon}) \nabla_x \phi. \tag{I.15}$$

Taking the time derivative again of (I.14) and the divergence of (I.15) and subtracting the resulting two equations leads to:

$$\partial_t^2 (n_i - n_e) - \nabla_x^2 : (S_i - S_e) = \nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} \right) \nabla_x \phi \right), \tag{I.16}$$

where ∇_x^2 denotes the tensor of second order derivatives and ":" the contracted product of two tensors. After substitution of the Poisson equation (I.7), this equation yields:

$$-\lambda^2 \partial_t^2 \Delta_x \phi - \nabla_x^2 : (S_i - S_e) = \nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} \right) \nabla_x \phi \right). \tag{I.17}$$

Collecting ϕ into the left-hand side, we find the so-called reformulated Poisson equation:

$$-\nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} + \lambda^2 \partial_t^2 \right) \nabla_x \phi \right) = \nabla_x^2 : (S_i - S_e). \tag{I.18}$$

These computations show that, if the triple (f_e, f_i, ϕ) solves the Vlasov-Poisson problem (I.5)-(I.7), it solves the "reformulated Vlasov-Poisson problem" consisting of the Vlasov equations (I.5), (I.6) and the reformulated Poisson equation (I.18). Conversely, if the triple (f_e, f_i, ϕ) solves the reformulated Vlasov-Poisson problem, we find, going backwards in the previous computations, that ϕ satisfies:

$$\partial_t^2 (-\lambda^2 \Delta_x \phi - (n_i - n_e)) = 0.$$

Therefore, if the initial data $\phi_0 := \phi|_{t=0}$ and $\phi_0' := (\partial_t \phi)|_{t=0}$ satisfy the two Poisson equations at the initial time:

$$-\lambda^2 \Delta_x \phi_0 = (n_i - n_e)_0, \tag{I.19}$$
$$-\lambda^2 \Delta_x \phi_0' = (n_i - n_e)_0' = -\nabla_x \cdot ((nu)_i - (nu)_e)_0, \tag{I.20}$$

(with obvious notations), then the Poisson equation (I.7) is satisfied at all times. Since eq. (I.18) is second order in time, it requires the knowledge of the two initial conditions ϕ_0 and ϕ_0'. Therefore, it is always possible to impose (I.19) and (I.20). A linearized stability analysis would in principle be needed to guarantee that, when (I.19) and (I.20) are not satisfied (for instance, in the discrete case, because of numerical errors), unstable linear modes are not going to be excited. However, linearized stability analysis depends on the state about which the problem is linearized and can only offer definite answers in model cases. The numerical experiments so far have not shown any instability problem of this kind (see section 4).

We can summarize this discussion by saying that the Vlasov-Poisson system (I.5)-(I.7) is **equivalent** to the following **"reformulated Vlasov-Poisson system"**:

$$\partial_t f_i + v \cdot \nabla_x f_i - \nabla_x \phi \cdot \nabla_v f_i = 0, \tag{I.21}$$

$$\partial_t f_e + v \cdot \nabla_x f_e + \frac{1}{\varepsilon} \nabla_x \phi \cdot \nabla_v f_e = 0, \tag{I.22}$$

$$-\nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} + \lambda^2 \partial_t^2 \right) \nabla_x \phi \right) = \nabla_x^2 : (S_i - S_e), \tag{I.23}$$

with ϕ satisfying the initial conditions (I.19), (I.20).

The reformulated Poisson equation has been previously proposed in the framework of fluid models in [10, 12, 14], and in [2, 13] for plasma kinetic models.

2.3 The quasineutral limit

The quasineutral limit of the Vlasov-Poisson system has been investigated theoretically in [5, 20, 21] and extensively used in physical studies. It is almost impossible to cite all the relevant physical literature. We just mention [26, 29, 39] as examples.

The reformulation (I.21)-(I.23) allows to investigate the quasineutral limit $\lambda \to 0$ in a straightforward way. Indeed, letting $\lambda \to 0$ in (I.23), (I.19), (I.20) directly provides the following equation for the quasineutral potential ϕ:

$$-\nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} \right) \nabla_x \phi \right) = \nabla_x^2 : (S_i - S_e), \tag{I.24}$$

together with the two constraints

$$(n_i - n_e)_0 = 0, \tag{I.25}$$

$$(n_i - n_e)_0' = -\nabla_x \cdot ((nu)_i - (nu)_e)_0 = 0, \tag{I.26}$$

>From these equations and with the aid of (I.16), we immediately deduce that $n_i = n_e$ at all times, which shows that quasineutrality holds.

Therefore, the limit $\lambda \to 0$ of the Vlasov-Poisson system (I.5)-(I.7), or equivalently, of its reformulation (I.21)-(I.23) leads to the following **quasineutral Vlasov system**:

$$\partial_t f_i + v \cdot \nabla_x f_i - \nabla_x \phi \cdot \nabla_v f_i = 0, \tag{I.27}$$

$$\partial_t f_e + v \cdot \nabla_x f_e + \frac{1}{\varepsilon} \nabla_x \phi \cdot \nabla_v f_e = 0, \tag{I.28}$$

$$-\nabla_x \cdot \left(\left(n_i + \frac{n_e}{\varepsilon} \right) \nabla_x \phi \right) = \nabla_x^2 : (S_i - S_e), \tag{I.29}$$

together with the initial conditions satisfying (I.25), (I.26).

We see that, although the original and reformulated Vlasov-Poisson systems are equivalent, they are not equally well-suited in the quasineutral limit. Indeed, the quasineutral potential eq. (I.29) appears as the formal limit $\lambda \to 0$ of the reformulated Poisson eq. (I.23) but not of the original Poisson eq. (I.7). The asymptotics does not preserve the form of the original Poisson equation while it does preserve the form of the reformulated one. Therefore, the construction of Asymptotic-Preserving schemes must be based on the use of the reformulated Poisson equation. For this reason, our Asymptotic-Preserving PIC method relies on the numerical approximation of the reformulated Vlasov-Poisson problem (I.21)-(I.23) rather than that of the original Vlasov-Poisson problem (I.5)-(I.7). We now describe the method in detail in the next section.

3 Asymptotic-Preserving PIC method (PICAP method)

3.1 PIC methods: general methodology

Particle-in-Cell (PIC) methods are widely used in the plasma physics community. We refer to the text books [3, 27] for a detailed exposition (see also the bibliography in section 1). The particle method consists in discretizing $f_{i,e}$ into a sum of delta measures located at positions $((X_{i,e})_j(t), (V_{i,e})_j(t))$ in phase space. This is written:

$$\begin{aligned} f_{i,e}(x,v,t) &\approx (f_{i,e})_N(x,v,t) \\ &:= \sum_{j=1}^{N} (\omega_{i,e})_j \, \delta\left(x - (X_{i,e})_j(t)\right) \, \delta\left(v - (V_{i,e})_j(t)\right) , \end{aligned} \tag{I.1}$$

where N is the number of particles and $(\omega_{i,e})_j$ is a weight which must be conveniently defined at initialization [3, 27]. The j-th particle coordinates $((X_{i,e})_j(t), (V_{i,e})_j(t))$ obey Newton's equations:

$$(\dot{X}_i)_j = (V_i)_j \, , \quad (\dot{V}_i)_j = -\nabla_x \phi_h((X_i)_j(t), t) \, , \tag{I.2}$$

$$(\dot{X}_e)_j = (V_e)_j \, , \quad (\dot{V}_e)_j = \frac{1}{\varepsilon} \nabla_x \phi_h((X_e)_j(t), t) \, , \tag{I.3}$$

These equations are discretized in time (see discussion below). The potential ϕ_h is a space approximation of ϕ. In the PIC method, this approximation is computed at each time step by solving the Poisson eq. (I.7) on a fixed grid of space step h (e.g. using a finite difference method). An assignment procedure allows to build grid values of the particle densities $(n_{i,e})_h$ from the knowledge of the particle locations and weights. These values serve as data for the numerical resolution of the Poisson equation. Once an approximation ϕ_h on the grid has been obtained, an interpolation procedure allows to reconstruct the values of the field $\nabla_x \phi_h(X_j, t)$ at the locations of the particles. These assignment-interpolation procedures are classical and are not modified in the present work (see again [3, 27]). The only modification resulting from the replacement of the original Poisson equation (I.7) by the reformulated one (I.23) is the need to assign other quantities than the densities, namely the values of the pressure tensors $S_{i,e}$. However,

the assignment procedure for these quantities is the same as for the densities and does not require a specific discussion.

The major issue in making the PIC procedure Asymptotic-Preserving, besides using the reformulated Poisson equation instead of the original one, is the time discretization. In the following subsections, we successively describe the “classical" or “standard" PIC method, the PICAP method, and for the sake of comparison, we also recall the Direct-Implicit PIC method.

3.2 Classical PIC method

The classical PIC method uses a “leap-frog" scheme where positions are defined at integer values of the time step $X_j^m \approx X_j(m\Delta t)$, while velocities are defined at half-integer values $V_j^{m+1/2} \approx V_j((m+1/2)\Delta t)$. In the classical PIC method, the integration of (I.2), (I.3) is done as follows:

$$\frac{(X_{i,e})_j^{m+1} - (X_{i,e})_j^m}{\Delta t} = (V_{i,e})_j^{m+1/2}, \tag{I.4}$$

$$\frac{(V_i)_j^{m+3/2} - (V_i)_j^{m+1/2}}{\Delta t} = -\nabla_x \phi^{m+1}((X_i)_j^{m+1}), \tag{I.5}$$

$$\frac{(V_e)_j^{m+3/2} - (V_e)_j^{m+1/2}}{\Delta t} = \frac{1}{\varepsilon}\nabla_x \phi^{m+1}((X_e)_j^{m+1}), \tag{I.6}$$

where we omit the index h for the grid-approximation of the potential for simplicity. Advancing the velocities with (I.5), (I.6) supposes that Poisson's equation (I.7) is solved with a charge density at the right-hand side of (I.7) computed from particle positions at time t^{m+1}, which is possible since these are known from (I.4). However, it is well known and widely documented in the literature [3, 27] that this method suffers from a stability constraint of the form $\Delta t, h \leq C\lambda$ where the constant C is of order of unity. We shall refer to this scheme as the classical PIC scheme. There have been numerous attempts to provide more stable time-stepping strategies (see the bibliography in section 1) which have proved to be quite efficient in practical cases. However, it is not clear if these methods are consistent with the quasineutral Vlasov model (I.27)-(I.29) in the limit $\lambda \to 0$. The PICAP methods which we are going to discuss now do have this property, as we will show below.

3.3 PICAP method

In the proposed Asymptotic Preserving strategy, positions and velocities are both defined at integer time-steps for simplicity. The time-stepping method for the position equation is fully implicit while that of the velocity equation is semi-implicit: the electric field is taken implicitly but the particle positions explicitly. More precisely, the particle

time advancement scheme is as follows:

$$\frac{(X_{i,e})_j^{m+1} - (X_{i,e})_j^m}{\Delta t} = (V_{i,e})_j^{m+1}, \tag{I.7}$$

$$\frac{(V_i)_j^{m+1} - (V_i)_j^m}{\Delta t} = -\nabla_x \phi^{m+1}((X_i)_j^m), \tag{I.8}$$

$$\frac{(V_e)_j^{m+1} - (V_e)_j^m}{\Delta t} = \frac{1}{\varepsilon}\nabla_x \phi^{m+1}((X_e)_j^m). \tag{I.9}$$

The implicit evaluation of the potential ϕ^{m+1} is obtained via an implicit time-discretization of the reformulated Poisson equation (I.23).

The starting point for our implicit time-discretization of (I.23) is the following scheme:

$$-\nabla_x \cdot \left((\Delta t)^2(n_i^m + \frac{n_e^m}{\varepsilon})\nabla_x \phi^{m+1} + \lambda^2(\nabla_x \phi^{m+1} - 2\nabla_x \phi^m + \nabla_x \phi^{m-1})\right) = \\ = (\Delta t)^2 \nabla_x^2 : (S_i^m - S_e^m). \tag{I.10}$$

This scheme is clearly Asymptotic-Preserving in the sense of the definition stated in section 1. Indeed, in the limit $\lambda \to 0$, we find

$$-\nabla_x \cdot \left((n_i^m + \frac{n_e^m}{\varepsilon})\nabla_x \phi^{m+1}\right) = \nabla_x^2 : (S_i^m - S_e^m). \tag{I.11}$$

which is a consistent discretization of the quasineutral potential eq. (I.29).

However, the presence of the Laplacians of the potential at the previous time steps ϕ^m and ϕ^{m-1} at the left-hand side of (I.10) is somehow inconvenient because it introduces extra computational cost. Therefore, it is more natural to use the Poisson eq. (I.7) to replace these terms by the charge densities at the corresponding time steps.

To this aim, two strategies have been proposed. The first method, already proposed in [13] and called "PICAP-1", consists in simply performing this replacement. It leads to:

$$-\nabla_x \cdot \left(((\Delta t)^2(n_i^m + \frac{n_e^m}{\varepsilon}) + \lambda^2)\nabla_x \phi^{m+1}\right) \\ = (\Delta t)^2 \nabla_x^2 : (S_i^m - S_e^m) + 2(n_i^m - n_e^m) - (n_i^{m-1} - n_e^{m-1}). \tag{I.12}$$

Eq. (I.12) allows to compute ϕ^{m+1} from known quantities at time t^m and t^{m-1}.

Another, original method is found by using a time-discrete version of the continuity eq. (I.14) to replace the term $(n_i^m - n_e^m) - (n_i^{m-1} - n_e^{m-1})$ at the right-hand side of (I.12) by $-\nabla_x((nu)_i^m - (nu)_e^m)$. This leads to the so-called "PICAP-2" method:

$$-\nabla_x \cdot \left(((\Delta t)^2(n_i^m + \frac{n_e^m}{\varepsilon}) + \lambda^2)\nabla_x \phi^{m+1}\right) \\ = (\Delta t)^2 \nabla_x^2 : (S_i^m - S_e^m) + (n_i^m - n_e^m) - \Delta t(\nabla_x((nu)_i^m - (nu)_e^m)), \tag{I.13}$$

which allows to compute ϕ^{m+1} from known quantities at time t^m alone. This equation can be viewed as a discretization of the first order differential system (I.14)-(I.15). Indeed,

an implicit discretization of this system is given by:

$$\frac{(n_i^{m+1} - n_e^{m+1}) - (n_i^m - n_e^m)}{\Delta t} + \nabla_x \cdot \Big((nu)_i^{m+1} - (nu)_e^{m+1})\Big) = 0, \tag{I.14}$$

$$\frac{((nu)_i^{m+1} - (nu)_e^{m+1})) - ((nu)_i^m - (nu)_e^m))}{\Delta t} + \nabla_x \cdot (S_i^m - S_e^m)$$
$$= -(n_i^m + \frac{n_e^m}{\varepsilon})\nabla_x \phi^{m+1}. \tag{I.15}$$

Inserting (I.15) in (I.14) leads to:

$$\frac{(n_i^{m+1} - n_e^{m+1}) - (n_i^m - n_e^m)}{\Delta t} + \nabla_x \cdot ((nu)_i^m - (nu)_e^m))$$
$$-(\Delta t)\nabla_x^2 : (S_i^m - S_e^m) - (\Delta t)\nabla_x \cdot \Big((n_i^m + \frac{n_e^m}{\varepsilon})\nabla_x \phi^{m+1}\Big) = 0. \tag{I.16}$$

Then replacing $(n_i^{m+1} - n_e^{m+1})$ by $-\lambda^2 \Delta \phi^{m+1}$ leads to (I.13).

The advantage of the PICAP-2 method is that it does not require an auxiliary scheme to compute the first iterate at time t^1, while PICAP-1 does. The computation of the first iterate requires the use of the classical PIC scheme. But if λ is very small, the use of the classical PIC scheme over a single time step can be enough to trigger an instability of the whole method. Therefore, the use of the PICAP-2 method is preferable.

In spite of the presence of the extra terms at the right-hand sides of (I.12) and (I.13) the PICAP-1 and PICAP-2 methods are Asymptotic-Preserving. Indeed, at each time step, both differ from the AP method (I.10) by terms which are proportional to $n_i - n_e$ and which consequently are of order $O(\lambda^2)$. Therefore, these additional terms vanish in the limit $\lambda \to 0$ and the PICAP-1 and PICAP-2 methods are Asymptotic-Preserving as well. The numerical experiments below provide an experimental evidence of this statement.

The space-discretization of the reformulated Poisson equations (I.12) or (I.13) uses standard finite-difference methods, and is omitted.

3.4 Direct-Implicit method

Here, for the sake of completeness, we describe the Direct-Implicit method of [31, 32] that we have implemented for our comparisons. To describe the Direct-Implicit method, we need to introduce $(\bar{E}_{e,i})_j^m$ the electric field applied at time t^m to the j-th particle, indices 'e' or 'i' refering to electrons or ions. The method is then as follows: given the electric field $(\bar{E}_{e,i})_j^m$ at time t^m, temporary particle positions and velocities are computed with the "leap-frog" scheme,

$$(\widetilde{V}_e)_j^{m+\frac{1}{2}} = (V_e)_j^{m-\frac{1}{2}} - \Delta t \frac{1}{\varepsilon}((1-\beta)(\bar{E}_e)_j^m - \beta \nabla_x \widetilde{\phi}^{m+1}((X_e)_j^m)), \tag{I.17}$$

$$(\widetilde{V}_i)_j^{m+\frac{1}{2}} = (V_i)_j^{m-\frac{1}{2}} + \Delta t((1-\beta)(\bar{E}_e)_j^m - \beta \nabla_x \widetilde{\phi}^{m+1}((X_i)_j^m)), \tag{I.18}$$

$$(\widetilde{X}_{i,e})_j^{m+1} = (X_{i,e})_j^m + \Delta t (\widetilde{V}_{i,e})_j^{m+\frac{1}{2}}, \tag{I.19}$$

where β is a parameter which controls the degree of implicitness ($0 \le \beta \le 1$) and $\widetilde{\phi}^{m+1}$ is a temporary potential which will be defined below. Then, temporary densities $\widetilde{n}_{i,e}^{m+1}$ are

assigned from the particle positions $(\widetilde{X}_{i,e})_j^{m+1}$ and the electric field at time t^{m+1} is given by the following implicit correction to the explicit electric field:

$$(\bar{E}_{i,e})_j^{m+1} = (1-\beta)(\bar{E}_{i,e})_j^m - \beta\nabla_x\phi^{m+1}((\widetilde{X}_{i,e})_j^{m+1}), \tag{I.20}$$

where ϕ^{m+1} is a correction to $\widetilde{\phi}^{m+1}$ and is a solution to the equation:

$$\begin{aligned} -\nabla_x\cdot\left(\left((\Delta t)^2\beta(\widetilde{n}_i^{m+1}+\frac{\widetilde{n}_e^{m+1}}{\varepsilon})+\lambda^2\right)\nabla_x\phi^{m+1}\right) = \\ \widetilde{n}_i^{m+1}-\widetilde{n}_e^{m+1}-\nabla_x\cdot\left((\Delta t)^2\beta(\widetilde{n}_i^{m+1}+\frac{\widetilde{n}_e^{m+1}}{\varepsilon})\nabla_x\widetilde{\phi}^{m+1}\right). \end{aligned} \tag{I.21}$$

Finally, the particle positions and velocities are updated with the classical "leap-frog scheme" from the positions $(X_{i,e})_j^m$ and velocities $(V_{i,e})_j^m$ at time t^m, in the partly implicit electric field $(\bar{E}_{i,e})_j^{m+1}$.

The temporary potential $\widetilde{\phi}^{m+1}$ can be chosen in various ways. In the following numerical simulations, it is chosen equal to 0. An other possible choice is $\widetilde{\phi}^{m+1} = \phi^m$. Simulations using the latter have been carried out in the present work but did not give rise to significant improvements compared to the former (and even sometimes provided mildly unstable results). For this reason, only simulation results using $\widetilde{\phi}^{m+1} = 0$ will be shown.

In the following, we will consider the explicit-implicit Direct-Implicit method with $\beta = 1/2$: it will be called the Classical Direct-Implicit method and denoted C-DI($\beta = 1/2$). The first iterate of the scheme requires the computation of the initial potential via the classical Poisson equation (I.7) (to provide both the velocities $V_j^{1/2} \approx V_j(\Delta t/2)$ and the initial explicit electric field). As shown in the numerical simulations, instabilities can then be generated in under-resolved case. That is why we consider two modified Direct-Implicit schemes : Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) denote Direct-Implicit schemes where the initial potential is zero. Modified DI($\beta = 1/2$) is an explicit-implicit Direct-Implicit scheme while Modified DI($\beta = 1$) is a fully implicit Direct-Implicit scheme.

We note that (I.21) is consistent with the Poisson equation when $\Delta t \to 0$ for fixed λ. On the other hand, when $\lambda \to 0$ with fixed Δt, the resulting equation is

$$\begin{aligned} -\nabla_x\cdot\left((\Delta t)^2\beta(\widetilde{n}_i^{m+1}+\frac{\widetilde{n}_e^{m+1}}{\varepsilon})\nabla_x\phi^{m+1}\right) = \\ \widetilde{n}_i^{m+1}-\widetilde{n}_e^{m+1}-\nabla_x\cdot\left((\Delta t)^2\beta(\widetilde{n}_i^{m+1}+\frac{\widetilde{n}_e^{m+1}}{\varepsilon})\nabla_x\widetilde{\phi}^{m+1}\right). \end{aligned}$$

When $\Delta t \to 0$, this scheme leads to

$$\widetilde{n}_i^{m+1} - \widetilde{n}_e^{m+1} = 0, \tag{I.22}$$

which does not provide any equation for ϕ^{m+1}. Additionally, since $\widetilde{n}_i^{m+1}$ and $\widetilde{n}_e^{m+1}$ do not depend on ϕ^{m+1}, there is no way ϕ^{m+1} can be adjusted in such a way that the constraint (I.22) can be matched. In other words, the scheme (I.21) is not Asymptotic-Preserving in the sense of the definition stated in section 1.

We refer to [31, 32] for a more thorough presentation of the Direct-Implicit schemes and to Table I.1 for a summary of the definition of the different schemes.

Classical PIC	eqs. (I.4)-(I.5)-(I.6)	explicit
PICAP-1	Asymptotic Preserving Pic, eqs.(I.7)-(I.8)-(I.9)-(I.12) and $\phi^{-1} = 0$	implicit
PICAP-2	Asymptotic Preserving Pic, eqs.(I.7)-(I.8)-(I.9)-(I.13)	implicit
C-DI($\beta = 1/2$)	Classical Direct-Implicit, eqs. (I.17)-(I.18)-(I.19)-(I.20)-(I.21) with $\beta = 1/2$	explicit-implicit
M-DI($\beta = 1/2$)	Modified Direct-Implicit, eqs. (I.17)-(I.18)-(I.19)-(I.20)-(I.21) with $\beta = 1/2$ and $\phi^0 = 0$	explicit-implicit
M-DI($\beta = 1$)	Modified Direct-Implicit, eqs. (I.17)-(I.18)-(I.19)-(I.20)-(I.21) with $\beta = 1$ and $\phi^0 = 0$	implicit

Table I.1: Definition of the PIC schemes

4 Numerical results

In this section we show numerical results in one space dimension for the Vlasov-Poisson system. We simulate four test-cases and we compare the results obtained with the Classical PIC, PICAP-1 and PICAP-2 schemes. Simulations with the Direct-Implicit schemes, C-DI($\beta = 1/2$), M-DI($\beta = 1/2$) and M-DI($\beta = 1$), are also performed in most of the test-cases.

The first test-case consists in a steady state, the second test-case is a bump-on-tail instability, the third test-case is a perturbation of a Maxwellian plasma and the fourth test-case is a plasma expansion in vacuum.

4.1 Steady state test-case

The first test case is dedicated to the study of the damping resulting from the implicitness of the PICAP schemes in the case of a steady state and they are compared to the Direct-Implicit schemes. We initialize the Vlasov-Poisson equation with the following steady state: the ion are supposed motionless and the distribution function of electrons is uniform,

$$f_0(x, v) = (2\pi)^{-\frac{1}{2}} \exp(-v^2/2), \quad n_0 = 1, \tag{I.23}$$

on the domain $(0, 1)$. We consider periodic boundary conditions for the Vlasov equation and homogeneous Dirichlet boundary conditions for the Poisson equation. The Debye length is chosen equal to $\lambda = 10^{-3}$ and thus the plasma period is $\omega^{-1} = 10^{-3}$.

In Fig. I.1, we consider a resolved case: $\Delta x = 10^{-4} < \lambda$. The time step is chosen such as it satisfies the CFL conditions: $v_{\max}\Delta t = 0.9\Delta x$, where $v_{\max}$ is the maximal electron velocity at each time step. As a consequence, time scales are also resolved: $\omega\Delta t \leq 1$. We use 100 particles per cell. Fig. I.1 shows the total energy as function of time. For the exact solution of the Vlasov-Poisson equation (one-species case), the total energy $\mathcal{E}$

is given by

$$\mathcal{E} = \frac{\lambda^2}{2} \int |\partial_x \phi|^2 dx + \frac{1}{2} \int f |v|^2 dx dv, \tag{I.24}$$

and is constant in time. The Classical PIC and the Classical Direct-Implicit schemes preserve the total energy while the PICAP-1, PICAP-2 and M-DI($\beta = 1$) schemes damp it since they are fully implicit. The damping rates related to these schemes are reported in Table I.2, Minimizing this damping effect by combining explicit and implicit computations

PICAP-1	PICAP-2	MDI($\beta = 1$)
-27%	-20%	-6.5%

Table I.2: Damping rates

is the goal of the vast litterature about Direct-Implicit schemes [31, 32]. PICAP schemes seem more dissipative but further studies about second order time-steping strategies may improve this feature.

Fig. I.2 presents results for the under-resolved case $\Delta x = 10^{-1} > \lambda$, where $\lambda = 10^{-3}$ is unchanged. For Classical PIC and Classical DI($\beta = 1/2$) schemes, we use a uniform time step $\Delta t = 15\omega^{-1}$, since the CFL conditions would be too restrictive due to the large particle velocities induced by the instability. For the other schemes, the time step is given by the CFL condition $v_{\max} \Delta t = 0.9 \Delta x$. Fig. I.2 (right) shows that time is under-resolved for all the schemes. We consider now 10^5 particles per cell in order to have the same total number of particles as in the resolved case. Total energies are plotted in Fig. I.2 (left) for the different schemes (Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$)). We note that the Classical PIC and Classical DI($\beta = 1/2$) methods are unstable: unphysically large electrical energies are reached after the first time step. The other schemes (PICAP-1, PICAP-2, Modified DI($\beta = 1/2$), Modified DI($\beta = 1$)) damp the energy. The damping rate is approximately of the same order and is about - 0.01%. Since the number of particles per cell is larger than previously and the time step bigger, the dissipation due to the particle-grid assignement-interpolation procedure is lower.

4.2 Periodic perturbation of a quasineutral Maxwellian plasma

In this section, we investigate the dynamics of a small perturbation of a Maxwellian plasma (see also [13]). We perform the simulation on the domain $(0, 1)$ with periodic boundary conditions for the Vlasov system and with homogeneous Dirichlet boundary conditions for the Poisson equation.

We consider successively the one-species case (i.e. system (I.8)-(I.9)) and the two-species case (i.e. system (I.5)-(I.7)). In [13], only the one-species case was considered. In the two-species case, we choose a realistic mass ratio $\varepsilon = 10^{-4}$. The PICAP schemes seem robust even in the case of realistic electron to ion mass ratios. However, ideally, an Asymptotic-Preserving scheme for both the limits $\lambda \to 0$ and $\varepsilon \to 0$ should be sought. This problem is under current investigation.

4.2.1 Periodic perturbation: one species case

For this case, we develop the PICAP-1 and PICAP-2 schemes for system (I.8)-(I.9). We select $\lambda = 10^{-4}$ which means that the scaled plasma frequency has the value $\omega = \frac{1}{\lambda} = 10^4$ and we use 100 particles per cell. We initialize the Vlasov-Poisson equation with

$$f_0 = \pi^{-1/2}(1 + \delta \sin(\kappa\pi x)) \exp(-v^2), \qquad n_0 = 1. \tag{I.25}$$

where $\delta = 10^{-2}$ is the perturbation amplitude and $\kappa = 2220$. The value of κ is chosen such that $\kappa \sim \lambda^{-1}$ to ensure that the wavelength of the density perturbation is of the same order as the Debye length. In this case, the phase velocity of the wave (which is nearly ω_p/κ in physical units) is of the same order as the thermal velocity v_{th} (which in our dimensionless units, is unity). This situation corresponds to a strong particle-wave coupling [7, 30], since the thermal velocity roughly coincides with the location of the steepest slope of the velocity distribution function.

In Fig. I.3, we present results obtained with the classical PIC, PICAP-1 and PICAP-2 schemes when $\Delta x = \lambda = 10^{-4}$. The time step Δt satisfies both CFL conditions $v_{\max}\Delta t \leq \Delta x$ and $\omega\Delta t \leq 1$. In these conditions, the fast space and time scales (respectively the Debye length and plasma period) are both resolved. Fig. I.3 (left) gives the electric potential as a function of position at an instant $t = 10\,\omega^{-1} = 10^{-3}$. The electric potential is almost identical with the three schemes. The electric potential as a function of position is shown on Fig. I.3 (right) after a large number of plasma periods ($t = 2000\,\omega^{-1} = 0.2$). We can see that the amplitude of the plasma waves is of the same order of magnitude as previously when using the classical PIC scheme, while it has been strongly damped out with the PICAP-1 and PICAP-2 methods. This shows that the AP strategy damps out the energy of the plasma waves and allows to capture phenomena which occur on longer time scales. In other words, the AP discretization of the Vlasov-Poisson problem damps out the plasma waves. On the other hand, waves that exist in quasineutral conditions (such as ion-acoustic or Alfven waves) should not be altered by the AP procedure. Indeed, the simulations of the ion slab expansion (see section 4.4) show that the model gives a correct account of the expansion, which propagates at about the ions acoustic wave-speed.

Fig. I.4 now shows the results obtained when the fast space and time scales are both under-resolved, i.e. the space step is larger than the Debye length $\Delta x > \lambda$ and the time step is larger that the plasma period $\Delta t > \omega^{-1}$. We choose $\Delta x = 10^{-2}$ while $\lambda = 10^{-4}$ is unchanged. Meanwhile, the total number of particles is kept unchanged. For both PICAP-1 and PICAP-2 schemes, we use a time step determined by the CFL conditions: $v_{\max}\Delta t \leq \Delta x$. In Fig. I.6, we plot $\omega\Delta t$ as a function of time. We see that this constraint still allows Δt of the order of 30 times the plasma period ω^{-1}. Simultaneously, we use a uniform time step for the Classical PIC scheme, with $\Delta t = 30\omega^{-1}$. We cannot use the CFL condition because the instability generates very large particle velocities and enforcing the CFL condition would generate very small time steps.

Fig. I.4 depicts the electric potential as a function of space, at time $t = 0.2$. The left picture shows a result using the classical PIC scheme. The instability of the scheme is clearly visible since the amplitude of the potential oscillations are now of the order of ten times those of the initial potential (see Fig. I.3 for instance). With the PICAP-1 or PICAP-2 schemes, these amplitudes are now very small showing that the schemes are stable and have damped out plasma waves, as in the resolved case.

Fig. I.5 displays the total energy $\mathcal{E}$ (in log scale) as a function of time for the different schemes in the resolved and under-resolved situations. Fig. I.5 (left) shows the resolved case. We notice that the total energy for the classical PIC scheme decays slowly after a more rapid initial transient. By contrast, the total energy of both PICAP-1 and PICAP-2 schemes decays more steadily. In Fig. I.5 (right), the under-resolved case is considered. With the Classical PIC scheme, the total energy is unstable and reaches large and totally unrealistic values after a very rapid initial transient. With the AP schemes, they are both stable and damped to zero. We note that the PICAP-2 scheme seems to exhibit a slower energy decay than the PICAP-1 scheme which has been initially proposed in [13]. Paradoxically, the energy decay of the PICAP-1 and PICAP-2 seems far less pronounced in the under-resolved case than in the resolved case, probably because the number of assignement-interpolation procedures per unit time has been reduced, leading to a lower diffusive scheme.

In [13], the energy damping has been attributed to the noisy coefficients of the reformulated Poisson equations (I.12) (because the densities and the tensors S have to be computed from the particles and because additionally, a second order space derivative of S is needed as a source term in this equation). It has been noticed that, if high frequencies are cut off in order to eliminate the noise, the energy damping can be significantly reduced. The application of filtering techniques, or more generally, of noise reduction techniques to these new PICAP schemes will be the subject of future studies.

When the fast space scale is under-resolved but the fast time scale is resolved, the Classical PIC scheme is still unstable. We have performed simulations with $\Delta x = 10^{-2}$, $\lambda = 10^{-4}$ and $\Delta t = 0.9\omega^{-1}$. The results are very similar to those obtained in the fully (time and space) under-resolved case (and for this reason, the results are not displayed). This is to be compared to the behaviour of classical schemes for the fluid models (i.e. the Euler-Poisson problem in the quasineutral limit) [12], where it has been observed that the results are stable as long as the fast time scale is resolved, even if the fast space scale is under-resolved. It looks as if the kinetic problem was more unstable than the fluid one, since as soon as one of the fast time or space scale is under-resolved, the scheme becomes unstable. By contrast, both the PICAP-1 and PICAP-2 schemes are stable. The results are very similar to those of the fully resolved case and are not represented.

4.2.2 Periodic perturbation: two species case

In this section, we consider the two species problem (I.5)-(I.7) and choose the following parameters:

$$\varepsilon = 10^{-4}, \qquad \lambda = 10^{-4} \qquad \left(\omega = \frac{1}{\lambda\sqrt{\varepsilon}} = 10^{6}\right).$$

Again, we put 100 ions and 100 electrons in each cell. We initialize this test-case with

$$\begin{aligned} f_{e0} &= \varepsilon\, \pi^{-1/2}(1 + \delta \sin(\kappa\pi x)) \exp(-\varepsilon v^2), \\ f_{i0} &= \pi^{-1/2}(1 + \delta \sin(\kappa\pi x)) \exp(-v^2). \end{aligned}$$

with $\delta = 10^{-2}$ and $\kappa = 2220$.

We successively consider the case where the Debye length and electron plasma period are both resolved ($\Delta x = \lambda$, $\Delta t < \omega^{-1}$, Δt satisfying the CFL condition $v_{\max}\Delta t \leq \Delta x$), and where none of the Debye length and plasma period are resolved. In the latter case,

the time-step is chosen to satisfy the CFL condition: $\Delta t = 0.9\Delta x/v_{\max}$. Fig. I.8 shows the evolution of the time step as a function of time. We notice that the time step reaches values which are of the order of 25 times the plasma period. For the standard PIC scheme, the time step is kept fixed at this value for the same reason as explained above. We observe that the PICAP methods damp plasma waves in a similar way as in the one-species cases, and that they provide stable computations in the under-resolved case, while, in the latter case, the classical PIC is unstable. We omit the corresponding pictures as they are similar to Figs. I.3 and I.4. We only display the total energy as a function of time.

In the two-species case, the total energy of the Vlasov-Poisson problem is given by

$$\mathcal{E} = \frac{\lambda^2}{2}\int |\partial_x \phi|^2 dx + \frac{1}{2}\int f_i |v|^2 dxdv + \frac{\varepsilon}{2}\int f_e |v|^2 dxdv, \tag{I.26}$$

and is constant in time. The total energy (in log scales) as a function of time is displayed in Fig. I.7, in the resolved case (left) and under-resolved one (right). The conclusions are the same as in the one-species case: the energy dissipation of the Classical PIC scheme is lower than that of the two AP schemes in the resolved case. However, in the under-resolved case, the Classical PIC scheme exhibits a severe instability while the two AP schemes are still stable. Additionally, the energy dissipations of the AP schemes in the under-resolved case are slightly less pronounced than in the resolved case. In this two-species case, the two AP schemes seem to behave similarly as regards the energy dissipation properties and the lower energy dissipation of PICAP-2 compared with PICAP-1 is less apparent.

These results confirm that, in a situation where standard PIC methods would be unstable, the PICAP schemes remain stable and dissipate the electric energy of the plasma waves.

4.2.3 Linear Landau damping test-case

In this section, the PICAP-1 and PICAP-2 schemes are compared to both the Direct-Implicit scheme [31, 32] and the Eulerian solver used in [2]. The frequency κ of the density perturbation is of order one. The target of such a test case is to measure the accuracy of the numerical schemes for capturing nonlinear Landau damping, which is a phenomenon occuring on the time scales of the plasma oscillations. So as to perform comparisons, we initialize the Vlasov-Poisson equation as in [2]:

$$f_0 = (2\pi)^{-\frac{1}{2}}(1 + \delta \sin(\kappa x)) \exp(-v^2/2), \quad n_0 = 1, \tag{I.27}$$

on the interval $(0, 2\pi/\kappa)$. The perturbation amplitude is taken equal to $\delta = 10^{-2}$ and $\kappa = 1$. We take $\Delta x = 2 \times 10^{-2}$ and we consider 10^4 particles per cell (in average). Thus, the total number of particles is of the same order as in the previous test case.

Fig. I.9 shows results for a resolved case $\lambda = 1 > \Delta x = 2 \times 10^{-2}$. The time step is chosen such that the CFL condition $v_{\max}\Delta t = 0.9\Delta x$ is satisfied. As a consequence, time scales are also resolved: $\omega \Delta t \leq 1$. Fig. I.9 (left) depicts the norm of the electric field (in log scale) as a function of time obtained with the classical PIC, PICAP-1, PICAP-2 and classical Direct-Implicit schemes. In Fig. I.9 (right), $\omega\Delta t$ as function of time is plotted. The four schemes give identical results. The measured slope is about 0.64, which has the same order as the theoretical one (see ref. [2], figure 3). However, due to the noise which

is inherent to particles simulations, this value is not as precise as for semi-Lagrangian simulations. For the same reason, the damping is stopped at $t \simeq 4\omega^{-1}$ instead of going on.

In Fig. I.10, we present the results for an under-resolved case: $\lambda = 10^{-4} < \Delta x$ and $\omega^{-1} = 10^{-4} < \Delta t$. The space step $\Delta x = 2 \times 10^{-2}$ is unchanged. For the classical PIC and the classical DI($\beta = 1/2$) schemes, we use a uniform time step $\Delta t = 30\omega^{-1}$. For the other schemes, the time step is determined by the CFL conditions: $v_{\max}\Delta t = 0.9\Delta x$. Fig. I.10 shows the norm of the electric field (in log scale) as a function of time for the different schemes. As in the previous test case, the classical PIC is unstable. The Direct-Implicit scheme seems also unstable but the stable results given by the modified DI($\beta = 1/2$) and modified DI($\beta = 1$) schemes prove that it is essentially due to the computation of the first time step. Finally, the modified DI($\beta = 1/2$) and the modified DI($\beta = 1$) schemes provide similar results to those of the PICAP-1 and PICAP-2 schemes. These results are not comparable to those obtained in [2]. This is because plasma oscillations are not resolved at all in this situation and none of the tested schemes is able to provide a reliable estimate of the damping rate. Both Direct-Implicit and PICAP methods provide an over-damping of the plasma waves. In Fig. I.10 (right), we plot $\omega\Delta t$ as a function of time. We see that the damping of the wave by the PICAP-1, the PICAP-2, the modified DI($\beta = 1/2$) and the modified DI($\beta = 1$) schemes results in increased time steps.

4.3 Bump-on-tail test-case

In this section, we compare the AP methods to the Classical PIC scheme and the Direct-Implicit method in the case of a bump-on-tail instability, which is a form of the two-stream instabilities. The results must be also compared with those obtained with an Eulerian code in [2]. We initialize the Vlasov-Poisson equation with

$$f_0(x, v) = f_1(v)(1 + \delta \cos(\kappa x)), \tag{I.28}$$

where the function f_1 is given by:

$$f_1(v) = C\left(\exp(-v^2/2) + \alpha \exp(-(v - v_d)^2/2v_t^2)\right), \tag{I.29}$$

where C is a renormalization constant. Periodic boundary conditions for the Vlasov-Poisson system and homogeneous Dirichlet boundary conditions for the Poisson equations are considered. The numerical parameters are $\alpha = 0.04$, $\kappa = 0.3$, $v_d = 4.5$, $v_t = 0.5$ and $\alpha = 2/9$, and are those chosen in [2]. The space domain is $(0, 20\pi)$ and we consider periodic boundary conditions.

Fig. I.11 (left) shows results for a very resolved case, where $\lambda = 1$ and $\Delta x = 2.10^{-3}$. We consider 10 particles per cell (in average). We consider also a CFL condition like $\Delta t = 0.9\Delta x/v_{\max}$, which consequently ensures the resolution of time scales. Electrostatic energy is plotted in Fig. I.11 (left) for Classical PIC, PICAP-1, PICAP-2 and Classical Direct-Implicit shemes. It should be compared with Fig. 10 of [2]. The various schemes are in very good agreement one with each other up to time $100\omega^{-1}$. Besides, up to time $50\omega^{-1}$, they are also in good agreement with the result obtained in [2]. After time $50\omega^{-1}$, the results are altered by the damping due to the numerical noise of PIC methods. However, all the schemes capture the dynamics well, despite the small number of particles per cell.

We consider now the following under-resolved case: $\Delta x = 3$ while $\lambda = 10^{-1}$. We consider 6.10^4 particles per cell (in average). For the Classical PIC and the Classical DI($\beta = 1/2$) schemes, we enforce the condition $\Delta t = 4\omega^{-1}$ to be sure that time is under-resolved. For the PICAP-1, the PICAP-2, the Modified DI($\beta = 1/2$) and the Modified DI($\beta = 1$) schemes, the time step is computed from the CFL conditions $v_{\max}\Delta t = 0.9\Delta x$. Fig. I.12 (right) shows that time scales are under-resolved for all these schemes since $\Delta t \simeq 4\omega^{-1}$ for all the schemes. In Fig. I.12 (left), electrostatic energy is plotted for Classical PIC, PICAP-1, PICAP-2, Classical Direct-Implicit, Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes: while Classical PIC exhibits a large instability before damping, Classical DI($\beta = 1/2$) and Modified DI($\beta = 1/2$) rapidly damp the energy but not as much as PICAP-1 and PICAP-2. Fig. I.13 displays the electron velocity distribution function for the different schemes at time $t = 0, 200, 300$ and $2000\omega^{-1}$. Classical PIC does not give coherent results: all the particles have large velocities. Classical DI($\beta = 1/2$) strongly heats up the system since the number of particles with small velocities decreases strongly. The Modified DI($\beta = 1/2$) scheme does not seem unstable. However, among the last four schemes, it gives rise to the larger number of particles with fast velocities (see at $v = 6$). On the contrary, the Modified DI($\beta = 1$) scheme seems to produce the system with the largest number of particles with small velocities (see at $v = 2$). The PICAP methods provide intermediate results. All these results are similar to those presented in Fig. 14 of [2]. Moreover, the results obtained with the PICAP-1, PICAP-2, Classical Direct-Implicit, Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes are in concordance with the resolved case considered in Fig. 12 of [2].

So as to compare the loss of energy due to the numerical schemes, Fig. I.11 (right) shows the total energy with the resolved Classical PIC scheme (with $\lambda = 10^{-1}$, $\Delta x = 5.10^{-2}$ and 10^3 particles per cell) and with the under-resolved other schemes (with $\lambda = 10^{-1}$, $\Delta x = 3$ and 6.10^4 particles per cell). As in the steady state test case, we can observe that fully implicit schemes (PICAP-1,PICAP-2 and Modified DI($\beta = 1$)) damp energy approximately at the same level and more rapidly than the explicit-implicit schemes, i.e. the Classical DI($\beta = 1/2$) and the Modified DI($\beta = 1/2$) schemes. Comparisons with Fig. 15 of [2] show that the damping due to the interpolation-assignement procedure of PIC methods seems to be more important than the damping due to implicitness.

4.4 One-dimensional plasma expansion test-case

4.4.1 Setting of the problem

We consider a one-dimensional plasma expansion problem which is described in [22, 23]. This is a two-species problem where the ions initially occupy a slab of thickness L, while the electrons are initialized by a Maxwell-Boltzmann equilibrium with a self-consistent potential. The test problem consists in observing the expansion of the ion slab.

The initial electron temperature is 1000 times higher than the initial ion temperature. The simulation box is $[0, A]$ and the ions are initialized in $[0, L/2]$ with $L/2 \ll A$ (for reasons of symmetry, only half of the domain is simulated and a symmetry axis is set at $x = 0$). The mass ratio is $\varepsilon = 1/1836$. We scale the energies according to the electron thermal energy, using $\bar{\phi} = e\phi/(k_B T_{e0})$, $v_0 = k_B T_{e0}/m_i$ in (I.4). The scaled Vlasov-Poisson

system is the same as before: (I.5)-(I.7). The boundary conditions for the potential are

$$\phi(0) = 0, \quad \partial_x \phi(A) = 0. \tag{I.30}$$

To enforce that $x = 0$ is an axis of symmetry, we assume specular reflection for the distribution function (i.e. particles are reinjected with reversed velocities), while the right boundary is a purely absorbing one, i.e. particles exiting the domain at $x = A$ are free to leave it while no particle is reinjected.

The initial electron density is defined by the Boltzmann relation

$$n_{e0} = n_0 \exp \phi_0, \tag{I.31}$$

while the initial ion density is such that

$$n_{i0} = \begin{cases} n_0 & \text{for} \quad 0 \le x \le L/2, \\ 0 & \text{for} \quad L/2 \le x \le A. \end{cases} \tag{I.32}$$

The initial potential ϕ_0 is obtained by solving the nonlinear Poisson equation (I.7) associated with the initial electron and ion densities (I.31), (I.32) and with the boundary conditions (I.30). The distribution functions are initialized by

$$f_{e0} = n_{e0} M_e(v), \quad f_{i0}(x, v) = n_{i0} M_i(v), \tag{I.33}$$

where the electron and ion Maxwellians are given by

$$M_e(v) = (\varepsilon/(2\pi))^{1/2} \exp(-\varepsilon v^2/2), \quad M_i(v) = (2\pi\eta)^{-1/2} \exp(-v^2/(2\eta)), \tag{I.34}$$

and $\eta = T_{i0}/T_{e0}$ is the ion to electron temperature ratio. In this test problem, $\eta = 10^{-3}$, and $L/2 = 20\lambda$.

In [23], the numerical parameters are chosen as follows. The simulation domain length is $A = 3 \times 10^4 \lambda$. The space step is $\Delta x = 0.2\lambda$ and there are 4×10^5 particles per cell. This makes a total number of 6×10^{10} particles, which exceeds our own computer resources. For this reason, we use a smaller simulation domain $A = 10^3 \lambda$ with the same $\Delta x = 0.2\lambda$ and 2000 particles per cell (1000 electrons and 1000 ions), which amounts to about 5×10^6 particles in total.

4.4.2 Simulation 1: time and space resolved case (reference case)

We first use the same time step as [23] i.e. $\Delta t = 0.05\omega^{-1}$, where $\omega = 1/(\sqrt{\varepsilon}\lambda)$ is the electron plasma frequency and we view the results at time $T = 30\omega_i^{-1}$ where $\omega_i = \sqrt{\varepsilon}\omega$ is the ion plasma frequency. These time and space steps obviously resolve the plasma period and Debye length.

The electric potential and electric field are shown on Fig. I.14 and should be compared with Fig. 9 (a) and (b) of [23]. We observe that the boundary of the ion slab is located at $\sim 140\lambda$ as in [23]. This shows that the AP method gives a correct account of ion-acoustic waves, since the expansion roughly takes place at the speed of this wave. This correct account is preserved in the under-resolved case as shown in sections 4.4.3 and 4.4.4 below. This observation suggests that the AP discretization does not modify waves

that are present in quasineutral conditions, such as ion-acoustic or Alfven waves, even in under-resolved conditions.

We note that the values reached by the two electric field peaks, close to the center of the foil ($E \sim 0.01E_0$) and at the boundary of the ion slab ($E \sim 0.03E_0$), measured in units of $E_0 = (n_0 k_B T/\varepsilon_0)^{1/2}$ are in very good agreement with [23]. The ion density distribution is shown on Fig. I.15 (left) and the electron velocity distribution, on Fig. I.15 (right). They must be compared respectively with Fig. 10 (a) and Fig. 3 (a) of [23]. Note that we display the space-integrated velocity distribution function rather than the distribution function at $x = 0$ as in [23] because otherwise, there are too few particles to make a significant statistics. The ion and electron mean velocities are shown in Fig. I.16 and should be compared with Fig. 10 (b) of [23]. The ion density and velocity show an excellent agreement with [23]. The electron velocity distribution also shows similar features as in [23]: a flat top, then a steep gradient between 0.25 and 0.5 in units of $2v_0^2$ and then an abrupt change to a flatter gradient around 0.5. The electron mean velocities are quite noisy due to the lack of significant statistics in the region beyond the boundary of the ion slab. This large noise level is also apparent in [23].

4.4.3 Simulation 2: time and space under-resolved case

We now turn to fully under-resolved (in both time and space) simulations. To this aim, the space-step is set to the value $\Delta x = 4\lambda$, and the time step, to the value $\Delta t = 3\omega^{-1}$. We present the electric potential and electric field on Figure I.18. We observe that the classical PIC method gives nonsense: the order of magnitude of the potential and of the electric field have nothing to do with the true solution. The PICAP methods on the other hand give fairly good agreements with the reference solution. We still observe the same qualitative features as in the resolved case, with a peak electric field located at the boundary of the ion slab, and a smaller but distinct peak close to the center of the foil. On the other hand, the ion expansion is slowed down: the boundary of the ion slab is located at $x \sim 100\lambda$ in both PICAP-1 and PICAP-2 methods respectively as can be seen on the the ion density profiles on Fig. I.17 (left). This reduction of the ion expansion can again be attributed to the lower precision of the method due to the large time and space steps (Δt is 60 times bigger than in the reference case, while Δx is 20 t times bigger). All the Direct-Implicit schemes give less accurate results than the PICAP-1 and PICAP-2 schemes since they suffer from a lack of consistency in the quasi-neutral regime. The "best" Direct-Implicit schemes, the C-DI($\beta = 1/2$) and the M-DI($\beta = 1/2$), locate the boundary of the ion slab at $x \sim 75\lambda$, while the fully implicit M-DI($\beta = 1$) locates this boundary at $x \sim 35\lambda$. We may expect an even better agreement of the PICAP schemes by the use of second order time discretizations. This point will ve studied in future work.

Despite the slight degradation of accuracy, the PICAP methods provide fairly good results even in highly under-resolved situations, while the classical PIC method fails completely, and the Direct-Implicit methods give poorer results. The use of much larger time and space steps leads to a considerable speed-up of the numerical simulation over fully-resolved PIC simulations which makes the PICAP methods attractive compromises between accuracy and computational efficiency.

The ion density as a function of position (in log scale) and the electron velocity distribution as a function of $v^2/(2v_0^2)$ obtained in these under-resolved conditions are

shown on Fig. I.17 (left and right respectively). Again, apart from the slower expansion of the slab, the ion density profiles given by the PICAP methods are quite good, with a sharp decrease at the boarder of the slab, while the classical PIC method gives an almost uniform ion density, indicating that the ion slab has totally dissolved. We note that the PICAP-2 method seems slightly more accurate than the PICAP-1 method, as the ion expansion is faster with the former than with the latter. Both PICAP methods give more accurate results than the Direct-Implicit schemes. A close look at the electron velocity distribution shows that the PICAP methods better reproduce the distribution of low energy electrons ($v^2/2v_0^2 \leq 0.5$) than the Direct-Implicit methods, while the latter provide a better approximation of the large energy electrons ($v^2/2v_0^2 > 0.5$). It seems that a better account of the low energy electrons is preferable to provide a good description of the overall dynamics. The ion and electron mean velocities are shown on Fig. I.19. With the PICAP methods, the general trend of the ion mean velocity is preserved. The maximal ion velocity (attained at the boundary of the ion slab) is smaller than in the reference case, which is again consistent with the slower expansion of the ion slab. With the Direct-Implicit methods, the ion velocity is scaled down further in accordance with the slower expansion velocity.

4.4.4 Simulation 3: time and space under-resolved with small number of particles

These simulations show that the PICAP methods are able to produce fairly accurate results at much lower cost than the classical PIC method and with a better accuracy than the Direct-Implicit method. To emphasize this point, we now show simulation results using much less particles. Specifically, using the same under-resolved time and space steps, we now initialize the simulation with 1000 particles per mesh for each species, like in the reference simulation. Since the mesh size is 20 times bigger than in the reference simulation, there are 20 times less particles, namely a total of about 2.5×10^5 particles. The results are displayed in Figs. I.20 to I.22. The electric potential and electric field show almost no difference with the previous simulation using 20 times more particles (see Fig. I.18). The boundary of the ion slab as appearing on the density profiles in Fig. I.21 (left), is located at the same position as in Fig. I.17 (left). This indicates that the speed of the ion expansion is not affected by the number of particles. Because of the smaller number of particles, the statistics of the electron distribution function as appearing on Fig. I.21 (right) degrades at large velocities. However, for low velocities, the results are similar to the previous case (see Fig. I.17 (right)). The ion and electron mean velocities (Fig. I.22) are also very similar to Fig. I.19. The use of much larger time and space steps, together with a much smaller number of particles, results in considerably faster simulations. Of course, the price to pay is a slight degradation of accuracy. However, it appears that the PICAP method provides an interesting compromise between accuracy and computational efficiency and this accuracy is also sharper than the Direc-Implicit methods.

4.4.5 Computational speed-up and improvements

To illustrate this statement, we now compare the CPU time necessary to achieve the same final time step, using the time and space resolved classical PIC method on the one

hand (Simulation 1), and the time and space under-resolved PICAP method on the other hand (Simulations 2 and 3). Table I.3 displays the results. We normalize the CPU time to units of length and time equal to λ and ω^{-1} respectively and to one particle. Usually, the CPU time is normalized by the time and length steps but since the goal of the PICAP method is to use large time and space steps, we rather normalize the CPU time to the physical reference units.

The computational speed-up which can be obtained by the use of the PICAP method in the present situation is about 50 per particle. This means that Simulation 2, which uses the same number of particles as Simulation 1, is 50 times faster, but Simulation 3, which uses 20 times less particles, is about 1000 times faster. Given the extremely good qualitative agreement, this huge computationally speed-up can be extremely interesting for 2D and 3D simulations. Additionally, since the lower number of particles does not affect significantly the quality of the results, Simulation 3 also represents a considerable saving in terms of memory storage, which is also extremely interesting for higher dimensional simulations.

	Simulation 1 Classical PIC fully resolved	Simulation 3 or 4 PICAP fully under-resolved	Speed-up per particle PICAP/Class. PIC
CPU (sec) per particle, per λ, per ω^{-1}	$2.88\ 10^{-7}$	$6.0\ 10^{-9}$	48

Table I.3: CPU Speed-up per particle. Note that a smaller number of particles can be used in conjunction with the PICAP method, thereby increasing the CPU speed-up (see text).

The degradation of accuracy related to the use of large time and space steps and small particle numbers seems to manifest itself in a slowing down of the time scales (such as that of the expansion of the ion slab). In general, in such circumstances, the method tends to overdamp the dynamics. However, the orders of magnitude are correct and the accuracy is significantly increased compared to Direct-Implicit methods. These tests also showed that there is no significant difference between the two PICAP methods, apart from a slightly lower dissipation by the PICAP-2 method. Additionally, the PICAP-2 method being a one-step time advancement method, it is easier to implement and should be preferred over PICAP-1.

In order to further increase the computational speed-up, one needs to bypass the limitation given by the CFL condition, namely that any given particle should not move more than one mesh size per time step. We have tried to loosen the CFL condition, but this rapidly leads to unacceptable errors. The reason is that the PICAP method is not asymptotic preserving with respect to the particle masses. Since the electron mass is the

smallest, the CFL constraint is more restrictive for the electrons. Therefore, it would be extremely useful to design an asymptotic preserving PIC method associated with the small electron to ion mass ratio. Such a method is not available yet. A less efficient but probably easier method would be to use a local time-step and to take advantage of the less severe CFL constraint in the regions where the particle velocities are small. The use of a local space-step would also allow large cells in the regions of smooth gradients. In these regions, the CFL constraint could be locally loosened. Finally, the accuracy of the method could be increased by designing a second-order time integration method for the particle trajectories. All these research directions show that there exists a large potential of improvements for this method. These ideas will be implemented in future work.

5 Conclusion

In this paper, we have presented a novel PIC method for the Vlasov-Poisson equation. The method is Asymptotic Preserving in the quasineutral limit, i.e. it is consistent with the quasineutral Vlasov equation in the limit of vanishing scaled Debye length. To validate this method, we have investigated several test problems: the first one consists of a steady state simulation; the second one considered a perturbation of a quasi neutral plasma with a particular consideration of the Landau damping; the third one is a bump-on-tail instability; the fourth one concerns the expansion of an ion slab in vacuum. All tests have confirmed that the method is stable even if the time and space steps are well above the values set by the plasma period and Debye length, while the standard PIC method is unstable in these conditions. The Direct-Implicit methods well capture plasma oscillations, but PICAP methods provide more consistant results for stiff test-cases such as the plasma expansion one. Despite the loss of accuracy associated with the use of large time and space steps, the method produces fairly accurate results and provides an attractive compromise between accuracy and computational efficiency. In the future, additional improvements of the method will be sought, such as for instance, removing the CFL constraint on the time step or finding second-order accurate time discretizations of the particle trajectories. The method will also be extended to electromagnetic plasma simulations through the coupling of the Vlasov equations to the Maxwell equations.

Bibliography

[1] T. D. Arber, R. G. L. Vann, A critical comparison of eulerian-grid-based Vlasov solvers, *J. Comput. Phys.*, 180(1) :339-357, 2002.

[2] R. Belaouar, N. Crouseilles, P. Degond, E. Sonnendrücker, An asymptotically stable semi-Lagrangian scheme in the quasi-neutral limit, *J. Sc. Comput.* 41(3) :341-365, 2009.

[3] C.K. Birdsall, A.B. Langdon, *Plasma physics via computer simulation*, Taylor & Francis, 2004.

[4] J. U. Brackbill, D. W. Forslund, An implicit method for electromagnetic plasma simulation in two dimensions, *J. Comput. Phys.*, 46(2) :271-308, 1982.

[5] Y. Brenier, Convergence of the Vlasov-Poisson system to the incompressible Euler equation, *Comm. PDE*, 25(3) :737-754, 2000.

[6] J. A. Carrillo, F. Vecil, Nonoscillatory interpolation methods applied to Vlasov-based models, *SIAM J. Sci. Comput.*, 29(3) :1179-1206, 2007.

[7] F. F. Chen, *Introduction to plasma physics and controlled fusion*, Plenum Press, 1974.

[8] B. I. Cohen, A. B. Langdon, A. Friedman, Implicit time integration for plasma simulation, *J. Comput. Phys.*, 46(1) :15-38, 1982.

[9] G.-H. Cottet, P.-A. Raviart, Particle methods for the one-dimensional Vlasov-Poisson equations, *SIAM J. Numer. Anal.*, 21(1) :52-76, 1984.

[10] P. Crispel, P. Degond, M.-H. Vignal, An asymptotically stable discretization for the Euler-Poisson system in the quasineutral limit, *C. R. Acad. Sci. Paris, Ser.I*, 341(5) :323-328, 2005.

[11] P. Crispel, P. Degond, M.-H. Vignal, Quasi-neutral fluid models for current-carrying plasmas, *J. Comput. Phys.*, 205(2) :408-438, 2005.

[12] P. Crispel, P. Degond, M.-H. Vignal, An asymptotic preserving scheme for the two-fluid Euler-Poisson model in the quasineutral limit, *J. Comput. Phys.*, 223 :204-234, 2007.

[13] P. Degond, F. Deluzet, L. Navoret, An asymptotically stable Particle-In-Cell (PIC) scheme for collisionless plasma simulations near quasineutrality, *C. R. Acad. Sci. Paris, Ser.I*, 343(9) :613-618, 2006.

[14] P. Degond, J.G. Liu, M.-H. Vignal, Analysis of an asymptotic preserving scheme for the Euler-Poisson system in the quasineutral limit, *SIAM Num. Anal.*, 46(3) :1298-1322, 2008.

[15] P. Degond, C. Parzani, M.-H. Vignal, Plasma expansion in vacuum: modeling the breakdown of quasineutrality, *SIAM Multiscale Modeling and Simulation*, 2(1) :158-178, 2003.

[16] F. Filbet, E. Sonnendrücker, Comparison of eulerian Vlasov solvers, *Computer Physics Communications*, 150(3) :247-266, 2003.

[17] A. Ghizzo, P. Bertrand, A Vlasov code for the numerical simulation of stimulated Raman scattering, *J. Comput. Phys.*, 87 :495, 1990.

[18] A. Ghizzo, P. Bertrand, M. Shoucri, E. Fijalkow, M. R. Feix, An eulerian code for the study of the drift-kinetic Vlasov equation, *J. Comput. Phys.*, 108(1) :105-121, 1993.

[19] K. Ganguly, H. D. Victory Jr. On the convergence of particles methods for multidimensional Vlasov-Poisson systems, *SIAM J. Numer. Anal.*, 26(2) :249-288, 1989.

[20] F. Golse, L. Saint-Raymond, The Vlasov-Poisson system with strong magnetic field in quasineutral regime, *Mathematical Models and Methods in Applied Sciences*, 13(5) :661-714, 2003.

[21] E. Grenier, Defect measures of the Vlasov-Poisson system in the quasineutral regime, *Comm. P.D.E.*, 20(7) :1189-1215, 1995.

[22] T. Grismayer, Etude théorique et numérique de l'expansion d'un plasma créé par laser : accélération d'ions haute énergie, PhD dissertation, Ecole Polytechnique, Palaiseau, France, December 2006.

[23] T. Grismayer, P. Mora, J. C. Adam , A. Héron, Electron kinetic effects in plasma expansion and ion acceleration, *Phys. Rev. E*, 77(6) :066407, 2008.

[24] S-Y. Ha, M. Slemrod, Global existence of plasma ion-sheaths and their dynamics, *Comm. Math. Phys.*, 238(1) :149-186, 2003.

[25] D. W. Hewett, A. B. Langdon, Electromagnetic direct implicit plasma simulation, *J. Comp. Phys.*, 72(1) :121-155, 1987.

[26] D. W. Hewett, C. W. Nielson, A Multidimensional Quasineutral Plasma Simulation Model, *J. Comp. Phys.*, 29(2) :219-236, 1978.

[27] R. W. Hockney, J.W.Eastwood, *Computer simulation using particles*, Institute of Physics, 1988.

[28] S. Jin, Efficient Asymptotic-Preserving (AP) schemes for some multiscale kinetic equations, *SIAM J. Sci. Comp.*, 21(2) :441-454, 1999.

[29] G. Joyce, M. Lampe, S. P. Slinker, W. M. Manheimer, Electrostatic Particle-in-Cell simulation technique for quasineutral plasma, *J. Comp. Phys.*, 138(2) :540-562, 1997.

[30] N.A. Krall, A.W. Trivelpiece, *Principles of plasma physics*, San Francisco Press, 1986.

[31] A. B. Langdon, B. I. Cohen, A. Friedman, Direct implicit large time-step particle simulation of plasmas, *J. Comp. Phys.*, 51(1) :107-138, 1983.

[32] A. B. Langdon, D. C. Barnes, Direct implicit plasma simulations, in: Multiple time scales, Academic Press, 1985.

[33] R. J. Mason, Implicit moment particle simulation of plasmas, *J. Comp. Phys.*, 41(2) :233-244, 1981.

[34] R. J. Mason, Implicit moment PIC-hybrid simulation of collisional plasmas, *J. Comp. Phys.*, 51(3) :484-501, 1983.

[35] R. J. Mason, Hybrid and collisional implicit plasma simulation models, in: Multiple time scales, Academic Press, 1985.

[36] R. J. Mason, An electromagnetic field algorithm for 2D implicit plasma simulation, *J. Comp. Phys.*, 71(2) :429-473, 1987.

[37] Y.-J. Peng, Asymptotic limits of one-dimensional hydrodynamic models for plasma and semi-conductors, *Chin. Ann. Math. B*, 23(1) :25-36, 2002.

[38] E. Pohn, M. Shoucri, G. Kamelander, Eulerian Vlasov codes, *Computer Physics Communications*, 166(2) :81-93, 2005.

[39] P. W. Rambo, Finite-grid instability in quasineutral hydrid simulations, *J. Comp. Phys.*, 118(1) :152-158, 1995.

[40] F. Valentini, P. Trávincek, F. Califano, P. Hellinger, A. Mangeney, A hybrid-Vlasov model based on the current advance method for the simulation of collisionless magnetized plasma, *J. Comput. Phys.*, 225(1) :753-770, 2007.

[41] J. M. Wallace, J. U. Brackbill, D. W. Forslund, An implicit moment electromagnetic plasma simulation in cylindrical coordinates, *J. Comp. Phys.*, 63(2) :434-457, 1986.

Figures

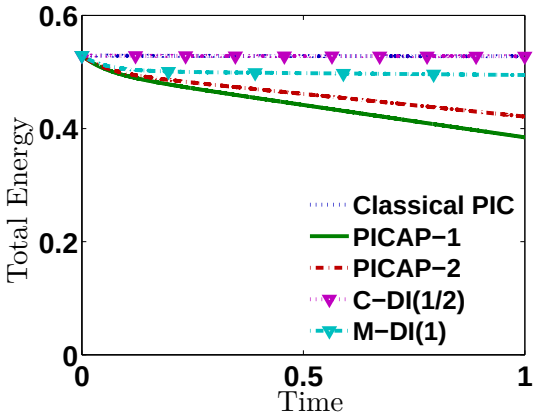


Figure I.1: Steady state in the resolved case: $\Delta x = 10^{-4} < \lambda = 10^{-3}$ and $\Delta t < \omega^{-1} = 10^{-3}$. Total energy as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes

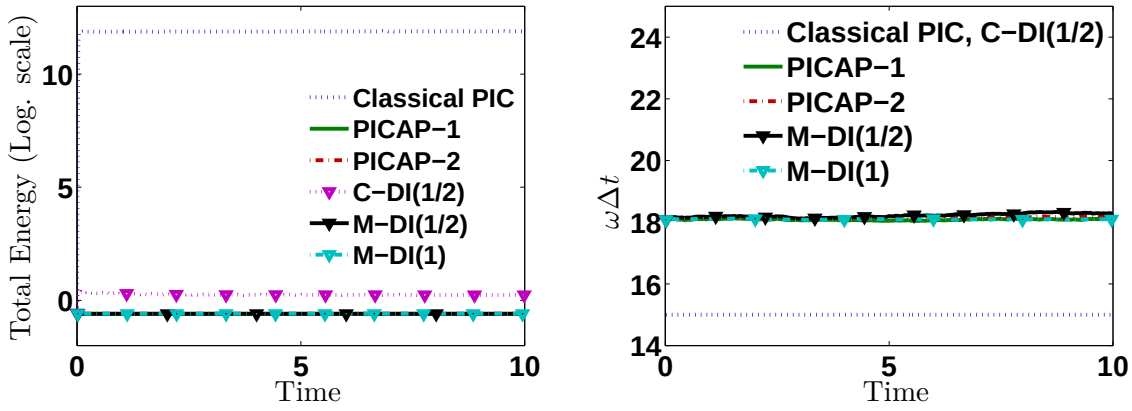


Figure I.2: Steady state in the under-resolved case: $\Delta x = 10^{-1} > \lambda = 10^{-3}$ and $\Delta t > \omega^{-1} = 10^{-3}$. Total energy in log scale (left) and $\omega\Delta t$ (right) as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes,

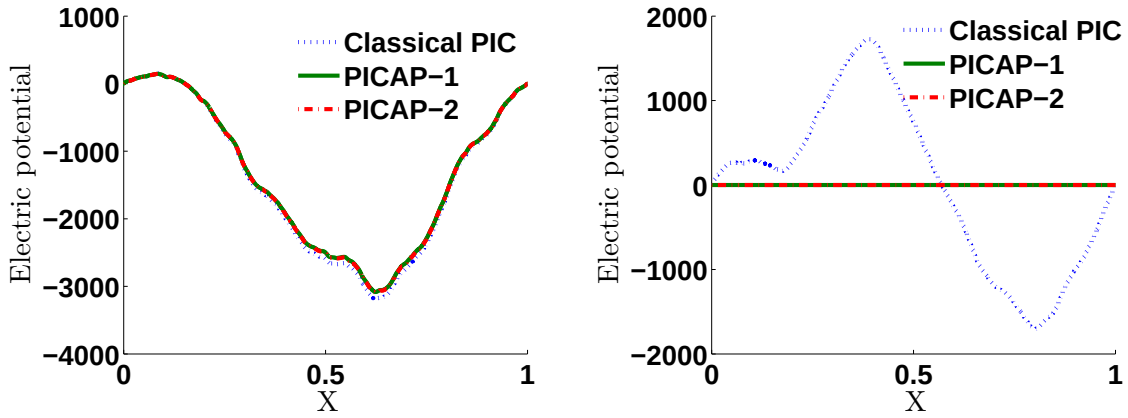


Figure I.3: One-species periodic perturbation of a quasi-neutral Maxwellian plasma, resolved case: $\Delta x = \lambda = 10^{-4}$ and $\Delta t < \omega^{-1} = 10^{-4}$. Electric potential as a function of position, with Classical PIC, PICAP-1 and PICAP-2 schemes. Left: at scaled time $t = 10\omega^{-1} = 10^{-3}$ (all curves are identical) ; right: at scaled time $t = 0.2 = 2000\omega^{-1}$.

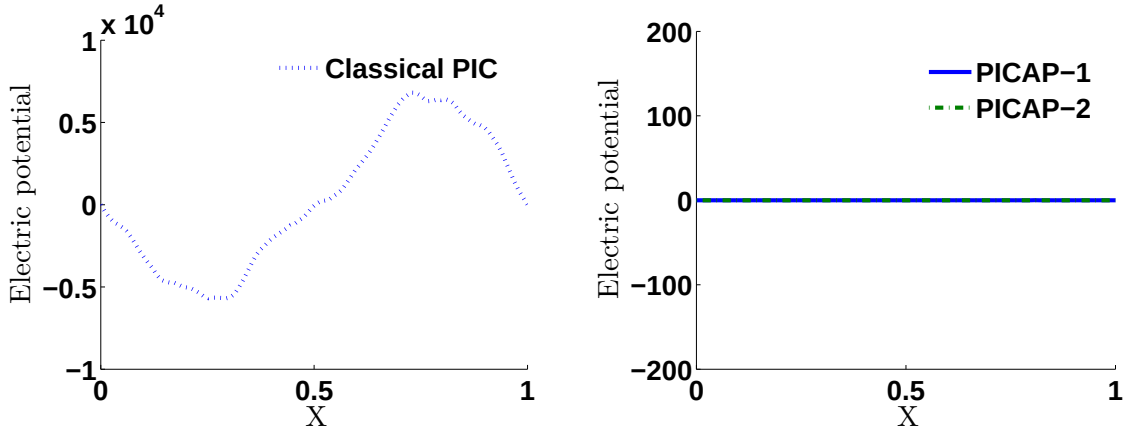


Figure I.4: One-species periodic perturbation of a quasi-neutral Maxwellian plasma, under-resolved case: $\Delta x = 10^{-2} > \lambda = 10^{-4}$ and $\Delta t > \omega^{-1} = 10^{-4}$. Electric potential as a function of position, with Classical PIC scheme (left), and PICAP-1, PICAP-2 schemes (right), at scaled time $t = 0.2 = 2000\omega^{-1}$.

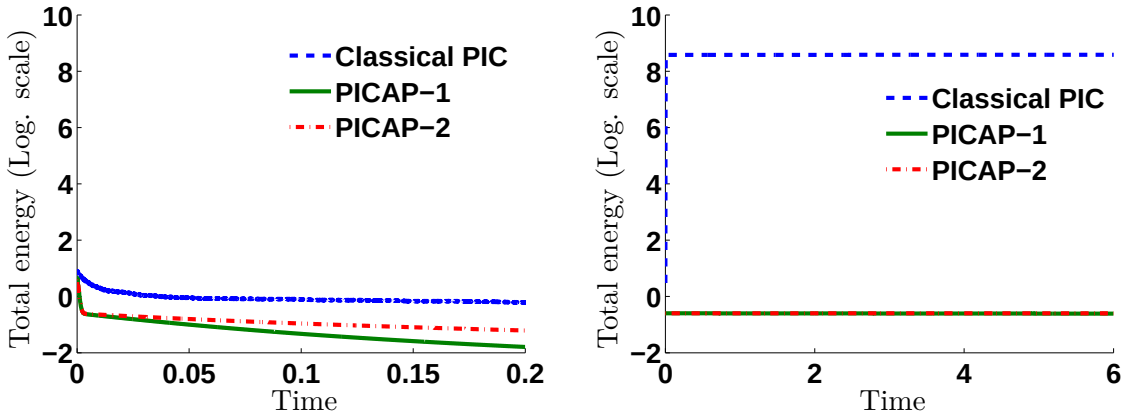


Figure I.5: One-species periodic perturbation of a quasi-neutral Maxwellian plasma, with $\lambda = 10^{-4}$, $\omega = 10^4$: total energy in log scale as a function of scaled time, with Classical PIC, PICAP-1 and PICAP-2 schemes. Left: resolved case: $\Delta x = \lambda$ and $\Delta t < \omega^{-1}$; right: under-resolved case: $\Delta x = 10^{-2} > \lambda$ and $\Delta t > \omega^{-1}$.

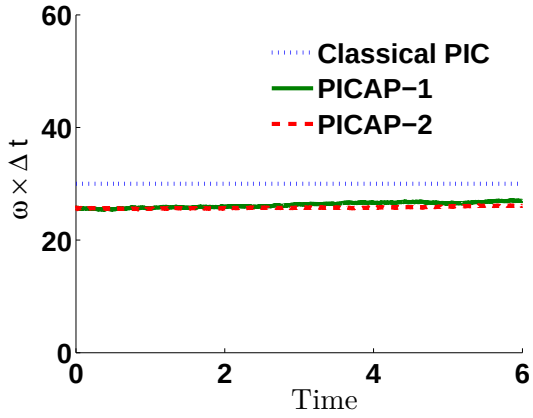


Figure I.6: One-species periodic perturbation of a quasi-neutral Maxwellian plasma, under-resolved case: $\Delta x = 10^{-2} > \lambda = 10^{-4}$ and $\Delta t > \omega^{-1} = 10^{-4}$. $\omega \Delta t$ as a function of scaled time, with Classical PIC, PICAP-1 and PICAP-2 schemes.

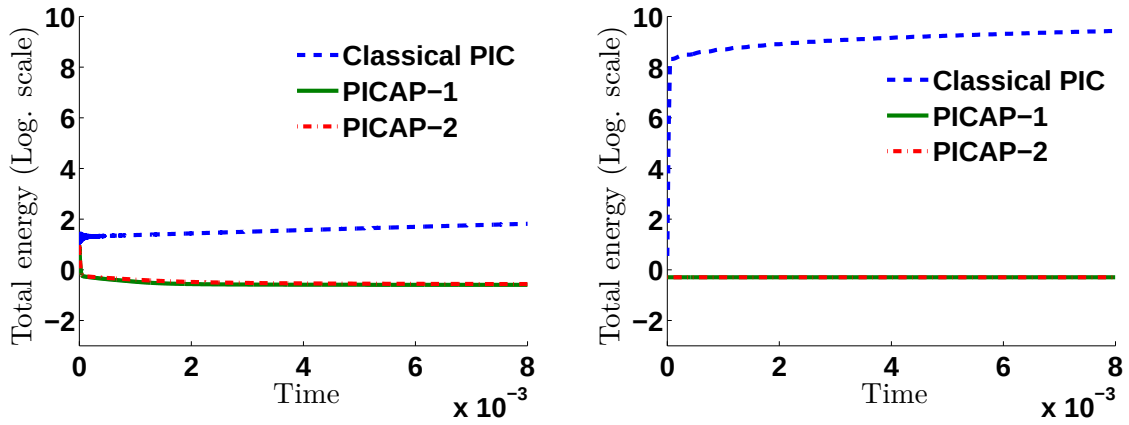


Figure I.7: Two-species periodic perturbation of a quasi-neutral Maxwellian plasma, $\varepsilon = 10^{-4}$, $\lambda = 10^{-4}$, $\omega = 10^6$: total energy (in log scales) as a function of scaled time, with Classical PIC scheme, PICAP-1 and PICAP-2 schemes. Left: resolved case: $\Delta x = \lambda$ and $\Delta t < \omega^{-1}$ ($\Delta t = 0.9\Delta x / v_{max}$) ; right: under-resolved case: $\Delta x = 10^{-2} > \lambda$ and $\Delta t > \omega^{-1}$ ($\Delta t = 0.9\Delta x / v_{max} \approx 25\omega^{-1}$).

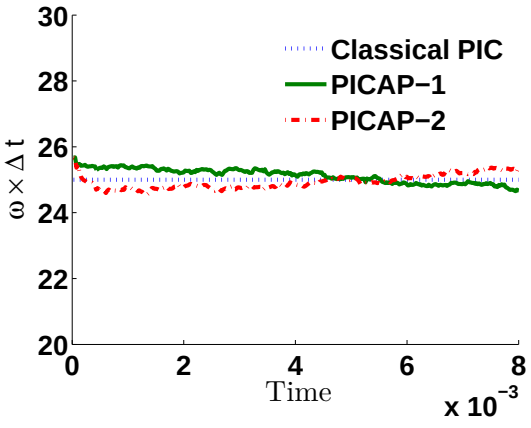


Figure I.8: Two-species periodic perturbation of a quasi-neutral Maxwellian plasma, under-resolved case: $\Delta x = 10^{-2} > \lambda$ and $\Delta t > \omega^{-1}$, $\varepsilon = 10^{-4}$, $\lambda = 10^{-4}$, $\omega = 10^6$. $\omega\Delta t$ as a function of scaled time, with Classical PIC, PICAP-1 and PICAP-2 schemes.

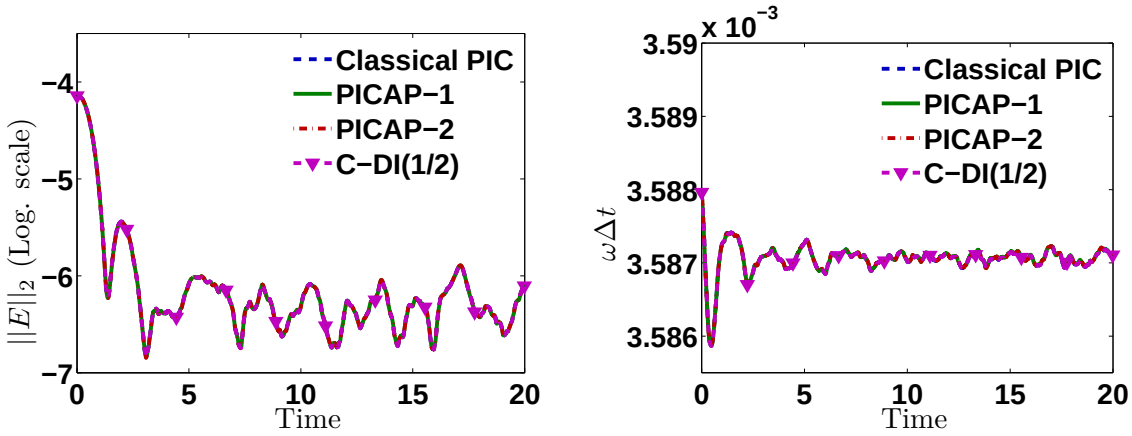


Figure I.9: Linear Landau damping in the resolved case: $\Delta x = 2 \times 10^{-2} < \lambda = 1$ and $\Delta t < \omega^{-1} = 1$. Left: norm of the electric field in log scale as a function of scaled time with Classical PIC, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$) schemes. Right: $\omega\Delta t$ as a function of scaled time, with Classical PIC, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$) schemes. This picture must be compared with Fig. 5 of reference [2].

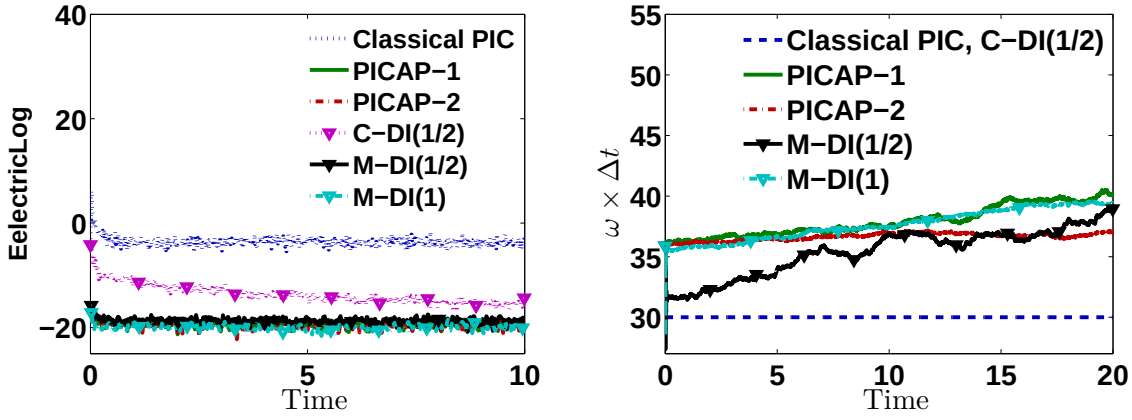


Figure I.10: Linear Landau damping in the under-resolved case: $\Delta x = 2 \times 10^{-2} > \lambda = 10^{-4}$ and $\Delta t > \omega^{-1} = 10^{-4}$. Left: norm of the electric field in log scale as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes. Right: $\omega\Delta t$ as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes. This picture must be compared with Fig. 17 of reference [2]

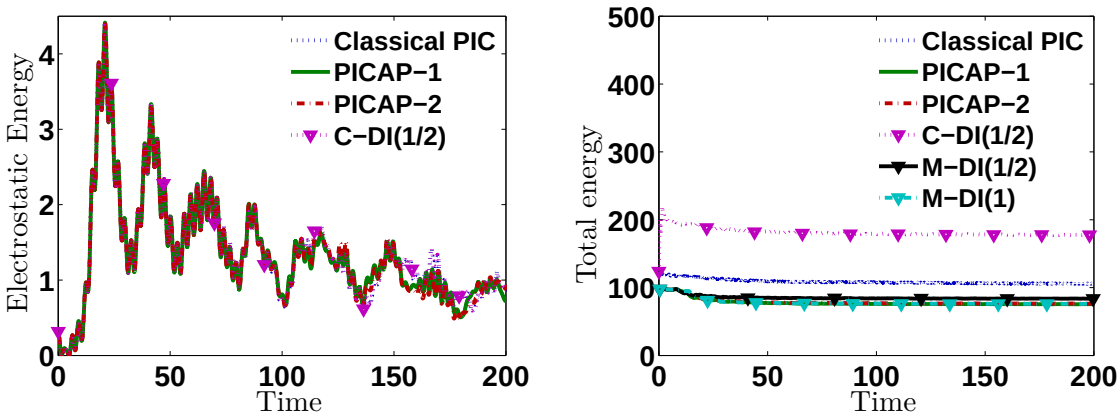


Figure I.11: Bump on tail instability. Left: electrostatic energy as a function of scaled time with Classical PIC, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$) schemes, in the resolved case: $\Delta x = 2 \times 10^{-3} < \lambda = 1$ and $\Delta t > \omega^{-1} = 1$. Right: comparison of methods: total energy as a function of scaled time, in the resolved case $\Delta x = 5 \times 10^{-2} < \lambda = 10^{-1}$ for classical PIC scheme and in the under-resolved case $\Delta x = 3 > \lambda = 10^{-1}$ for PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$), Modified DI($\beta = 1$) schemes. These pictures must be compared with Fig. 10 and Fig. 15 of reference [2].

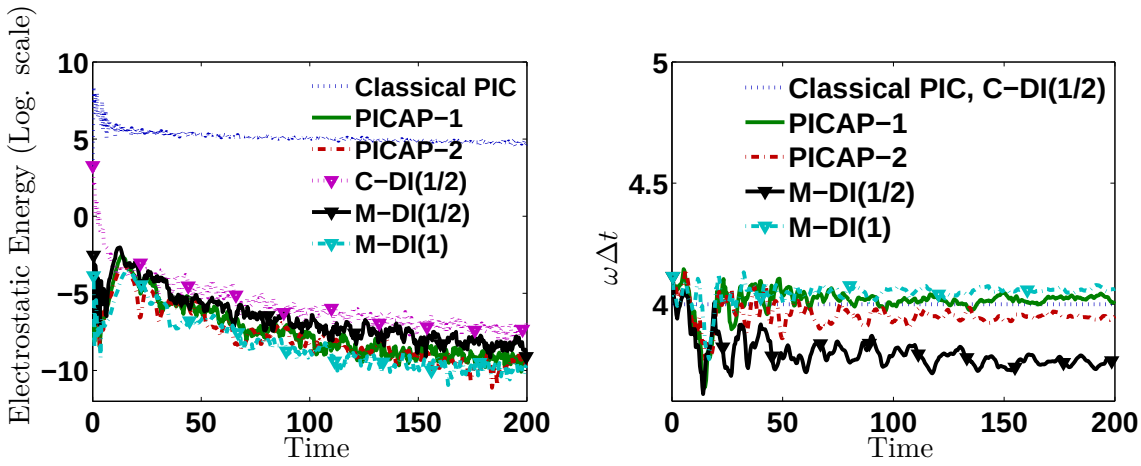


Figure I.12: Bump on tail instability. Right: electrostatic energy in log scale as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes, in the under-resolved case: $\Delta x = 3 > \lambda = 10^{-1}$ and $\Delta t > \omega^{-1} = 10^{-1}$. This picture must be compared with Fig. 13 of reference [2]. Left: $\omega \Delta t$ as a function of scaled time with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes, in the under-resolved case: $\Delta x = 3 > \lambda = 10^{-1}$ and $\Delta t > \omega^{-1} = 10^{-1}$.

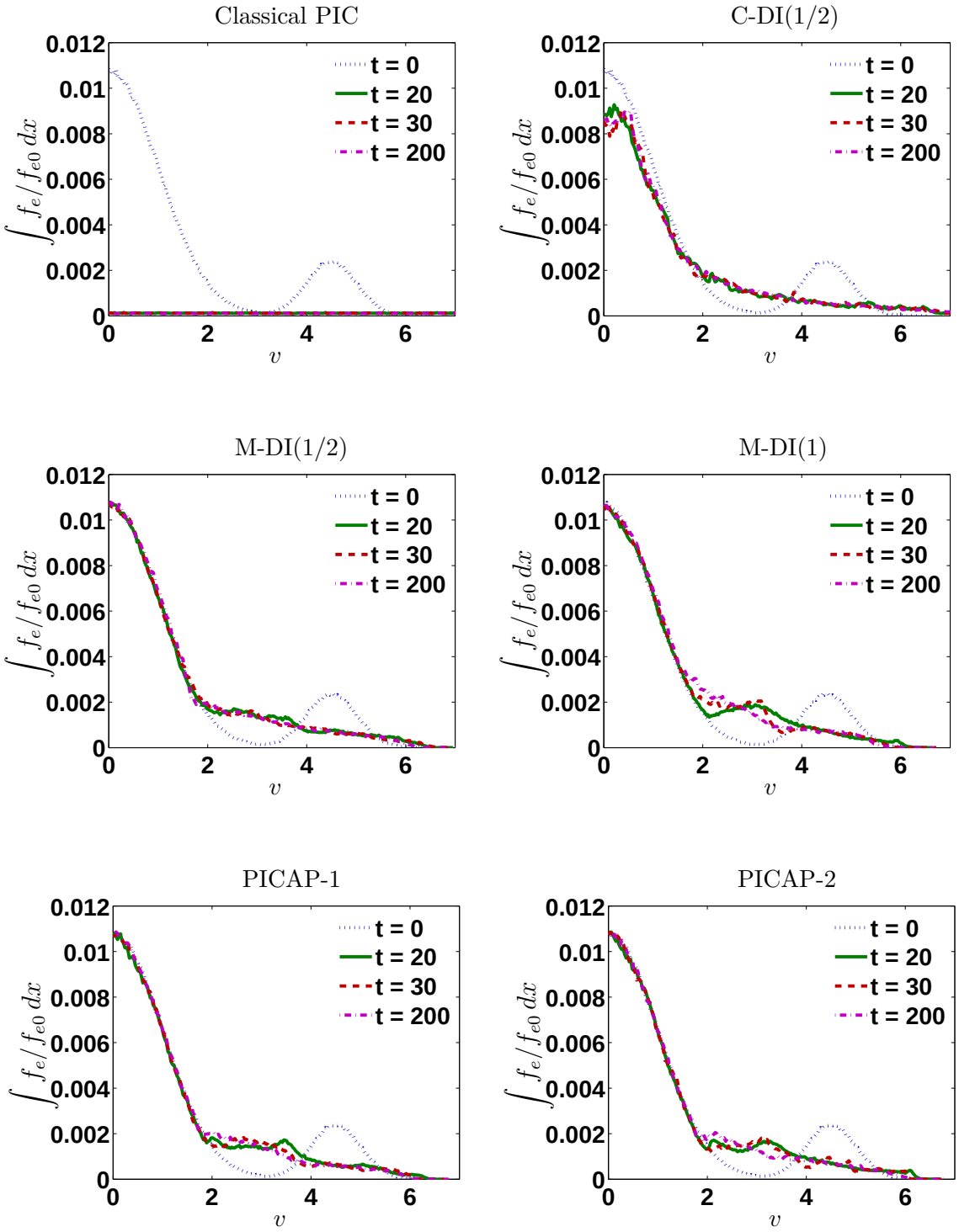


Figure I.13: Bump on tail instability in the under-resolved case: $\Delta x = 3 > \lambda = 10^{-1}$ and $\Delta t > \omega^{-1} = 10^{-1}$. Left top: velocity distribution function as a function of velocity with Classical PIC scheme. Right top: velocity distribution function as a function of velocity with Classical DI($\beta = 1/2$). Left medium: velocity distribution function as a function of velocity with Modified DI($\beta = 1/2$) scheme. Right top: velocity distribution function as a function of velocity with Modified DI($\beta = 1$) scheme. Bottom left: velocity distribution function as a function of velocity with PICAP-1 scheme. Bottom right: velocity distribution function as a function of velocity with PICAP-2 scheme.

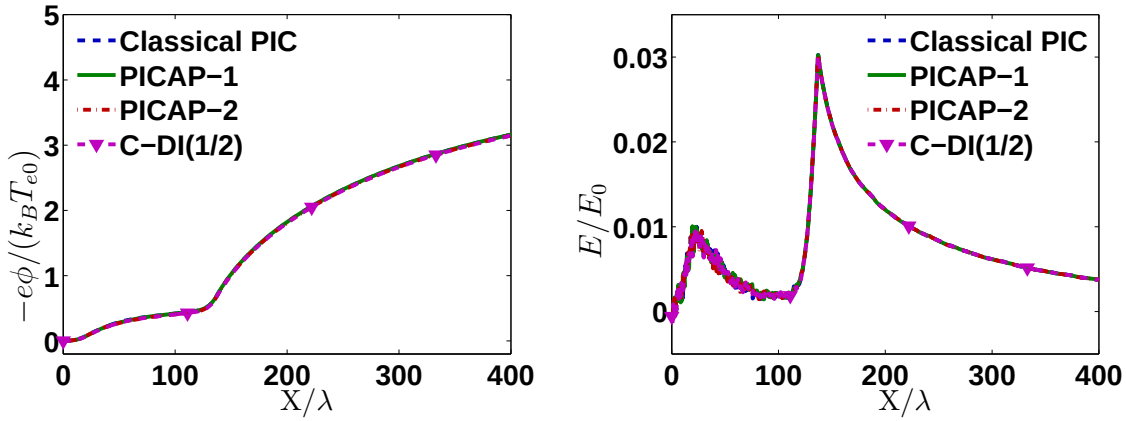


Figure I.14: Plasma expansion test case. Simulation 1: resolved case (reference case): $\Delta x = 0.2\lambda$ and $\Delta t = 0.05\omega^{-1}$. Electric potential (left) and electric field (right) as functions of position at time $t = 30\omega_i^{-1}$ with Classical PIC scheme, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$) schemes. These pictures must be compared with Fig 9 (a) and (b) of reference [23].

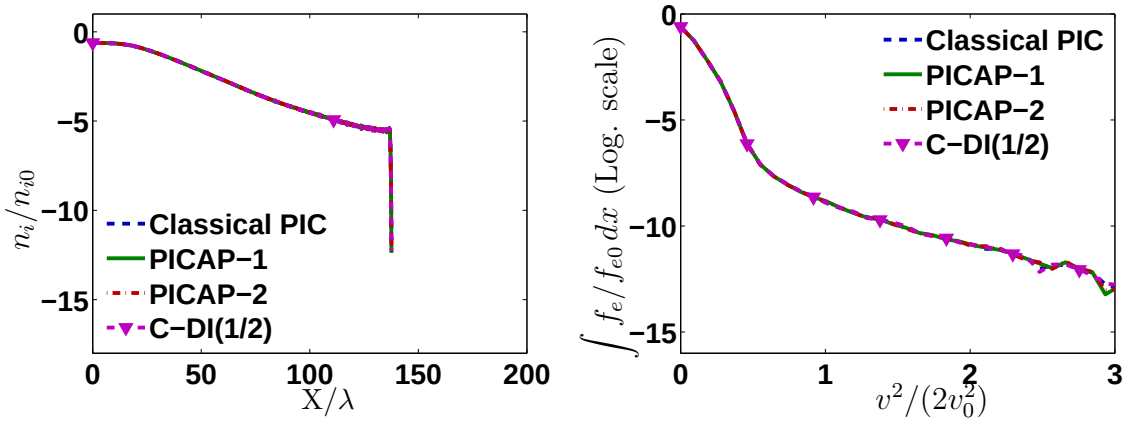


Figure I.15: Plasma expansion test case. Simulation 1: resolved case (reference case): $\Delta x = 0.2\lambda$ and $\Delta t = 0.05\omega^{-1}$. Ion density as a function of position in log scale (left) and electron velocity distribution as a function of $v^2/(2v_0^2)$ in log scale (right) with Classical PIC, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$), at time $t = 30\omega_i^{-1}$. These pictures must be compared with Fig 10 (a) and Fig. 3 (a) (respectively) of reference [23].

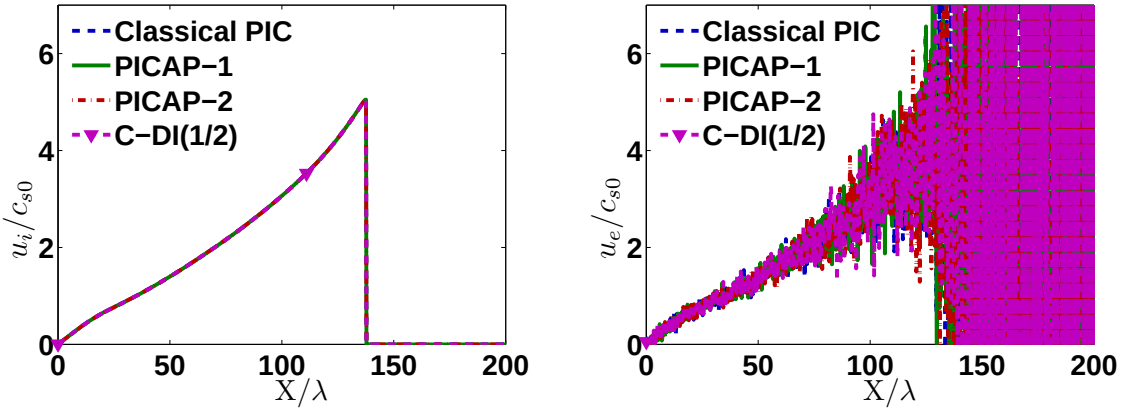


Figure I.16: Plasma expansion test case. Simulation 1: resolved case (reference case): $\Delta x = 0.2\lambda$ and $\Delta t = 0.05\omega^{-1}$. Ion and Electron mean velocities as functions of position, with Classical PIC, PICAP-1, PICAP-2 and Classical DI($\beta = 1/2$) schemes, at time $t = 30\omega_i^{-1}$ ($c_{s0} = \sqrt{k_B\, T_{e0}/m_i}$).

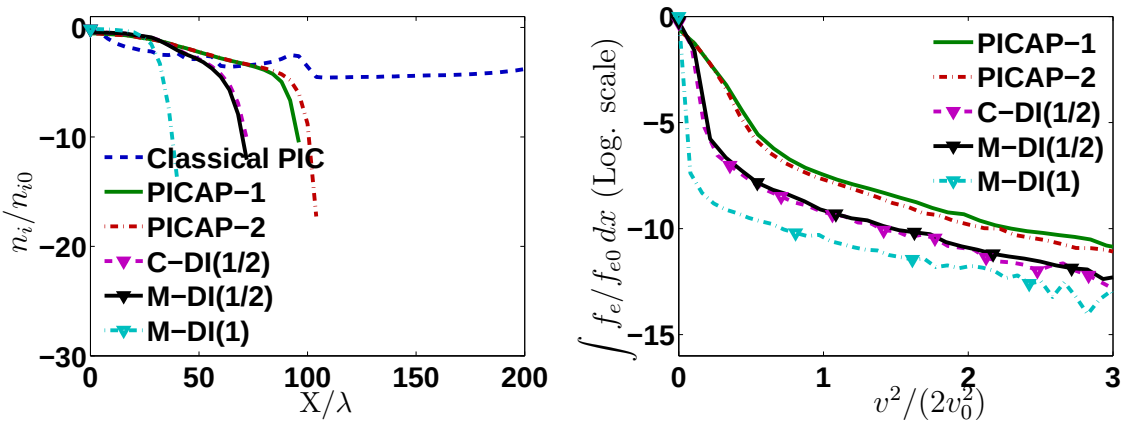


Figure I.17: Plasma expansion test case. Simulation 2: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$. Ion density as a function of position in log scale (left) and electron velocity distribution as a function of $v^2/(2v_0^2)$ in log scale (right) with classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes, at time $t = 30\omega_i^{-1}$.

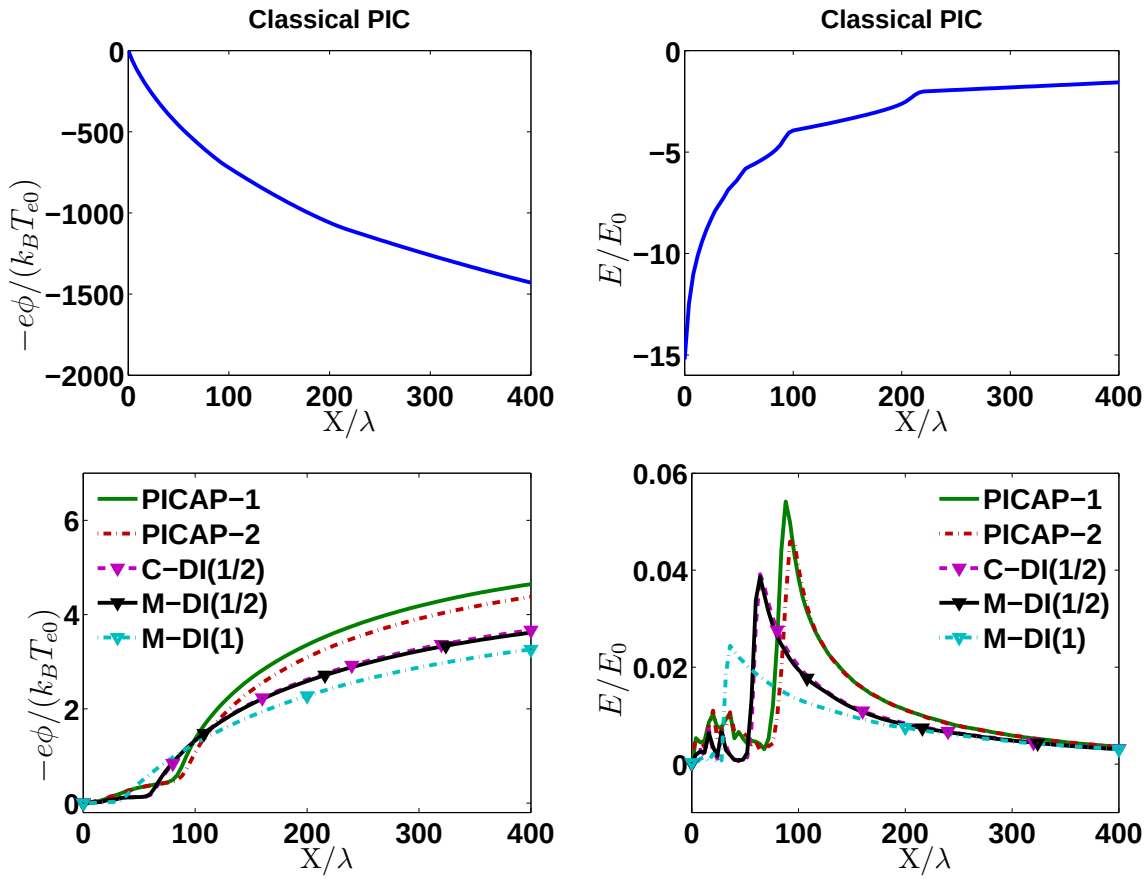


Figure I.18: Plasma expansion test case. Simulation 2: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$. Top: Electric potential (left) and Electric field (right) at time $t = 30\omega_i^{-1}$ as a function of position with classical PIC. Bottom: Electric potential (left) and Electric field (right) at time $t = 30\omega_i^{-1}$ as a function of position with PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes.

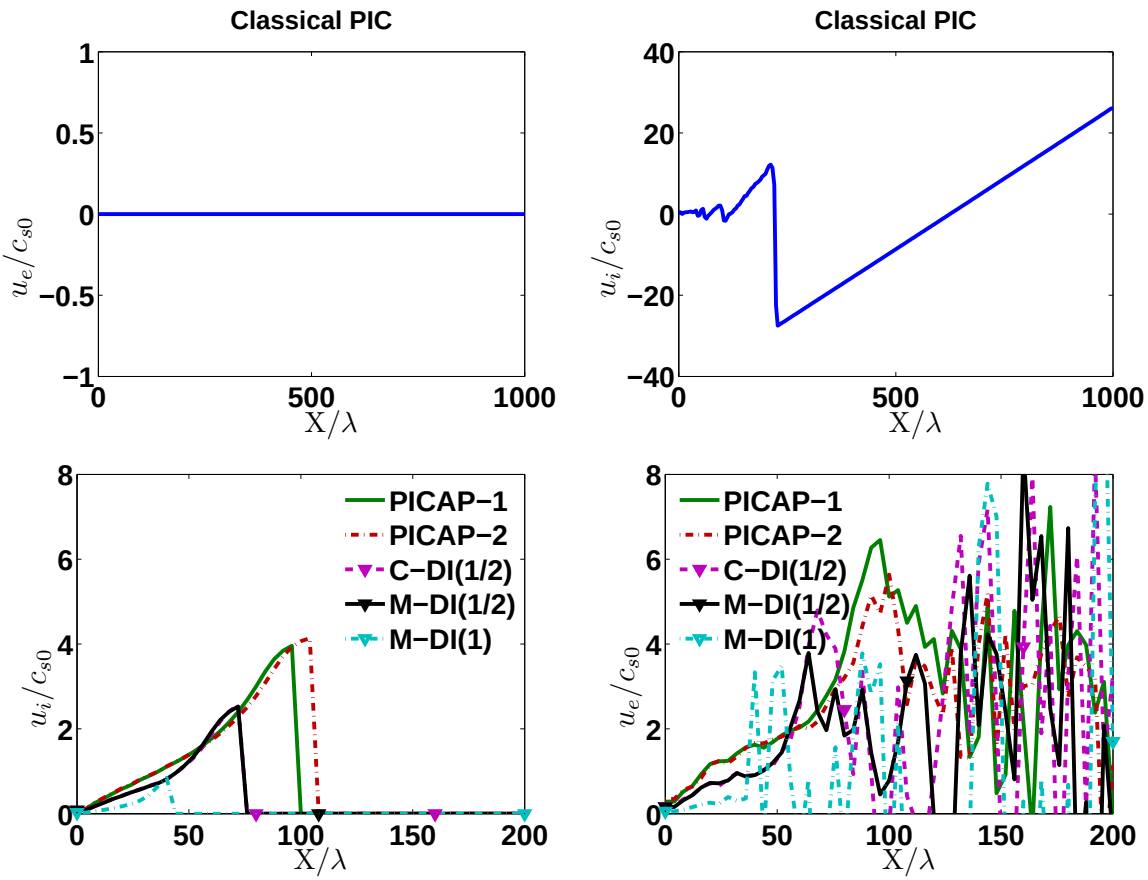


Figure I.19: Plasma expansion test case. Simulation 2: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$. Ion and Electron mean velocities as functions of position, with Classical PIC (top), PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes (bottom), at time $t = 30\omega_i^{-1}$ ($c_{s0} = \sqrt{k_B\, T_{e0}/m_i}$).

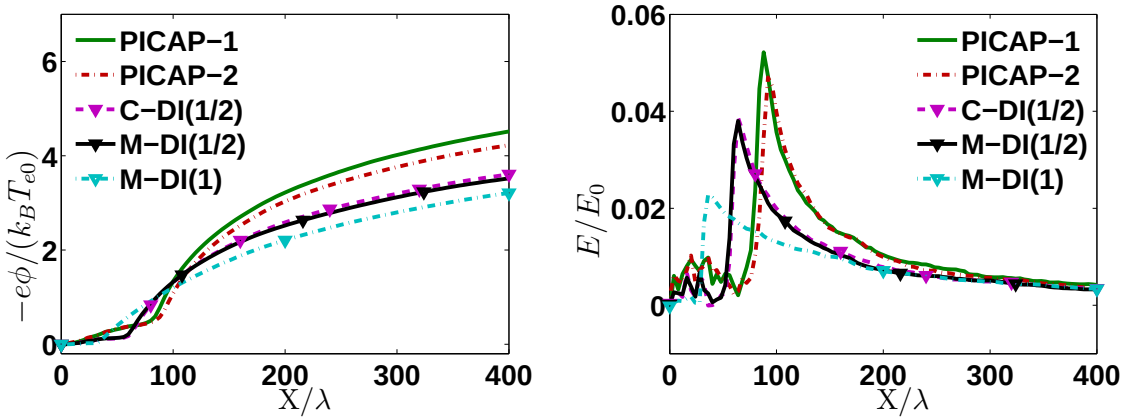


Figure I.20: Plasma expansion test case. Simulation 3: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$ and small number of particles (1000 particles per mesh and per species). Electric potential (left) and electric field (right) as a function of position with PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes at time $t = 30\omega_i^{-1}$.

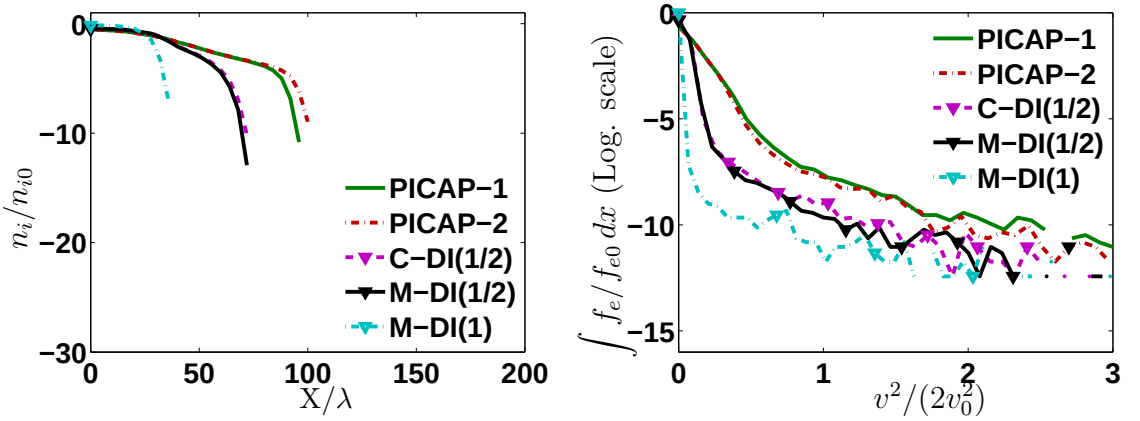


Figure I.21: Plasma expansion test case. Simulation 3: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$ and small number of particles (1000 particles per mesh and per species). Ion density as a function of position in log scale (left) and electron velocity distribution as a function of $v^2/(2v_0^2)$ in log scale (right) with Classical PIC, PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes, at time $t = 30\omega_i^{-1}$.

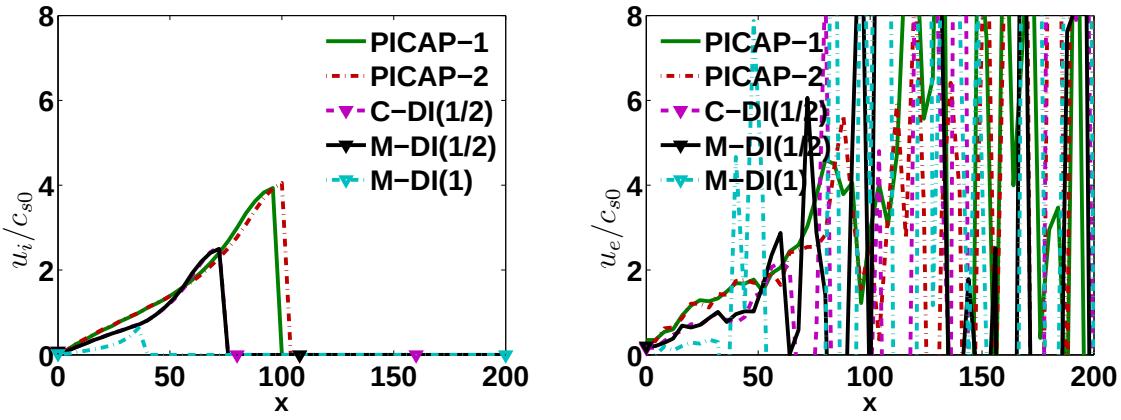


Figure I.22: Plasma expansion test case. Simulation 3: space and time under-resolved case: $\Delta x = 4\lambda$ and $\Delta t = 3\omega^{-1}$ and small number of particles (1000 particles per mesh and per species). Ion and electron mean velocities as functions of position, with PICAP-1, PICAP-2, Classical DI($\beta = 1/2$), Modified DI($\beta = 1/2$) and Modified DI($\beta = 1$) schemes, at time $t = 30\omega_i^{-1}$ ($c_{s0} = \sqrt{k_B\, T_{e0}/m_i}$).

Deuxième partie : Limite hydrodynamique avec contrainte géométrique

Chapitre II

Un modèle de Vicsek à deux populations

Ce chapitre est écrit en anglais.

A Vicsek model with two populations

1 Introduction

In this paper, we study a variant of the Vicsek model for the displacement of a system of particles which can either move or not. The Vicsek model, introduced in [4], is aimed at providing a minimal description of self-propelled particles with alignement interactions. In this original model, the particles move with a constant speed. A macroscopic version of the Vicsek dynamics was derived in [2]: the geometric constraint, satisfied by the velocities of the particles, implies that momentum is not conserved in time and thus the classical theory used for gaz [1] was generalized to such lack of conservation quantities. Here, we adapt the method when two populations with Vicsek interactions, one at rest (with zero speed but with non steady orientations) and an other moving (with speed one), are exchanging individuals depending on the local alignement of the two populations. We provide the dynamics at large time scales with respect to the scales of both Vicsek and exchange interactions.

This work is motivated by the behaviour of sheep inside the herds during grazing period: sheep alternate between motionless time when grazing and displacements to look for fresher grass. The change between the two states of the animal can be either spontaneous or triggered by fellows. Moreover, any member of the group can initiate a movement: this is called distributed leadership. This is one of the conclusions in [3], where biologists experimentally studied the following behaviour for gregarious sheep. In the model we consider, the interactions between moving and motionless sheep will be enhanced by their alignement. Experimental studies should be required to confirm or invalidate this assumption. In the following, we will study this model as a minimal model of synchronization between two populations with different characteristic speed. The comparison with experimental data is postponed to future works.

The plan of this chapter is as follows. In a first part, we introduce the two-phase model at the particle and kinetic level. We perform also a hydrodynamic rescaling so that all the interactions become local. In a second part, we investigate the asymptotic regime when the Vicsek interactions are more frequent than the exchanges between the two

populations. We obtain a macroscopic description for the densities and the macroscopic directions. In the third part, we briefly show that in the asymptotic regime when the exchanges between the two populations are more frequent than the Vicsek interactions, then the conditions to reach equilibria dynamics at large scale are more stringent than in the first asymptotic regime.

2 A Vicsek model with two speeds

2.1 The microscopic model

We consider N particles with positions $X_k \in \mathbb{R}^2$ and velocities $V_k = \eta_k \omega_k \in \mathbb{R}^2$ for $k \in \{1, \dots, N\}$, where $\eta_k \in \{0, 1\}$ and $\omega_k \in \mathbb{S}^1$ denote respectively the velocity moduli and the velocity directions. So, particles are separated into two groups: the group $G_0 = \{k,\ \eta_k = 0\}$ are particles at rest and the group $G_1 = \{k,\ \eta_k = 1\}$ are moving particles.

The interactions among particles of the same group are given by the Vicsek model, as described in [2]:

$$\begin{aligned}
&\frac{dX_k}{dt} = \eta_k \omega_k, \\
&d\omega_k = (\mathrm{Id} - \omega_k \otimes \omega_k)(\nu_{\eta_k} \bar{\omega}_k dt + \sqrt{2 d_{\eta_k}} dB_t), \\
&\bar{\omega}_k = \frac{J_k}{|J_k|}, \quad J_k = \sum_{j,\ \eta_j = \eta_k,\ |X_j - X_k| \le R} \omega_j,
\end{aligned}$$

where Id denotes the identity matrix, $w \otimes v$ denotes the tensor product of the two vectors w and v. The operator $(\mathrm{Id} - \omega_k \otimes \omega_k)$ is the projection operator onto the orthogonal plane to ω_k: it ensures the norm of ω_k to be unity. Two dynamics are in competition: particles are aligning with the mean direction $\bar{\omega}_k$ in the disc of radius R around them, and noise is applied to the direction with a white noise dB_t on $\mathbb{R}^2$. Alignment intensities, ν_0 and ν_1, and noise intensities, d_0 and d_1, are parameters controlling the relative strengh of the two dynamics for each group. Note that particles at rest are not moving in space but their directions are changing in time.

The particles can also change their speeds: it results in an exchange of particles between the moving and the unmoving populations. This exchange between the two populations are described by the dynamics of η_k: it is a time-continuous Markov process on the state space $\{0, 1\}$. The transition rates are given by:

$$g^k(\eta_k) = \frac{1}{\tau_{\eta_k}} \left(1 + \alpha \frac{1}{N} \sum_{j,\ \eta_j \neq \eta_k,\ |X_j - X_k| \le R} \frac{(1 + \omega_k \cdot \omega_j)}{2} \right); \tag{II.1}$$

$g^k(1)$ (resp. $g^k(0)$) is denoting the transition rate for a moving particle (resp. a particle at rest) to become a particle at rest (resp. a moving particle). The generator matrix is then given by:

$$G = \begin{pmatrix} -g^k(0) & g^k(0) \\ g^k(1) & -g^k(1) \end{pmatrix}.$$

The transition rates $g^k(1)$ and $g^k(0)$ consist of two parts: the first parts are intrinsic rates equal to $1/\tau_0$ and $1/\tau_1$ and the second parts are rates of order α/τ_0 and α/τ_1 depending on the local alignement with the members of the other group : for a particle at rest, the more aligned with the moving particles in its neighbourhood it is, the higher its probability to become moving is.

2.2 Kinetic model and scaling

Mean-field kinetic model. We introduce the two distribution functions in phase space: $f_0(x,\omega,t)$ for the particles at rest and $f_1(x,\omega,t)$ for the moving particles. With no exchange terms, they satisfy the Fokker-Planck-Kolmogoroff equations introduced in [2]:

$$\partial_t f_0 = \mathcal{Q}_0(f_0), \tag{II.2}$$
$$\partial_t f_1 + \omega\cdot\nabla_x f_1 = \mathcal{Q}_1(f_1), \tag{II.3}$$

where $\mathcal{Q}_0$ and $\mathcal{Q}_1$ are the Vicsek operators:

$$\mathcal{Q}_0(f_0) = -\nabla_\omega\cdot(\nu_0\mathcal{F}[f_0]f_0) + d_0\Delta_\omega f_0, \tag{II.4}$$
$$\mathcal{Q}_1(f_1) = -\nabla_\omega\cdot(\nu_1\mathcal{F}[f_1]f_1) + d_1\Delta_\omega f_1, \tag{II.5}$$

with $\nabla_\omega\cdot$ and Δ_ω are respectively the divergence and the laplacian operator on the circle $\mathbb{S}^1$. $\mathcal{F}$ denotes the alignement forces in each group and is given by:

$$\mathcal{F}[f](x,\omega,t) = (\mathrm{Id}-\omega\otimes\omega)\bar{\omega}(x,\omega,t),$$
$$\bar{\omega}(x,\omega,t) = \frac{\mathcal{J}(x,t)}{|\mathcal{J}(x,t)|},\quad \mathcal{J}(x,t) = \int_{y\in\mathbb{R}^2,v\in\mathbb{S}^1} K(|x-y|)vf(y,v,t)dydv.$$

where $K(|x|)$ is the interaction kernel equal to the indicator function of the disc of radius R. Without brownian motion ($d_0 = d_1 = 0$), the above equations are also satisfied by the empirical distribution functions and so the mean-field limit could be investigated. With brownian motion, no rigorous framework can justisfy the resulting equations at the kinetic level of the one-particle distribution function. Yet, recent numerical simulations provide a numerical validation at least in some range of parameters (see chapter III).

Similarly, we admit that the limit of the mean-field interactive particle system with the exchange dynamics is given by:

$$\partial_t f_0 = \mathcal{Q}_0(f_0) + \mathcal{E}(f_0,f_1), \tag{II.6}$$
$$\partial_t f_1 + \omega\cdot\nabla_x f_1 = \mathcal{Q}_1(f_1) - \mathcal{E}(f_0,f_1), \tag{II.7}$$

where the exchange term $\mathcal{E}(f_0,f_1)$ is given by:

$$\mathcal{E}(f_0,f_1) = -\frac{1}{\tau_0}\mathcal{G}[f_1]f_0 + \frac{1}{\tau_1}\mathcal{G}[f_0]f_1, \tag{II.8}$$
$$\mathcal{G}[f](x,\omega,t) = 1+\alpha\int_{y\in\mathbb{R}^2,v\in\mathbb{S}^1} K(|x-y|)\frac{(1+v\cdot\omega)}{2}f(y,v,t)dydv. \tag{II.9}$$

The system (II.6)-(II.7)-(II.8)-(II.9) is the starting point of our study.

Scaling of the interactions. Supposing that the Vicsek parameters ν_0, ν_1, d_0, d_1 are all of order $O(1/\delta_V)$ and the exchange parameters $1/\tau_0$, $1/\tau_1$ are of order $O(1/\delta_E)$, we can write equations (II.6)-(II.7) in a scaled form as following:

$$\partial_t f_0 = \frac{1}{\delta_V}\mathcal{Q}_0(f_0) + \frac{1}{\delta_E}\mathcal{E}(f_0, f_1), \tag{II.10}$$

$$\partial_t f_1 + \omega \cdot \nabla_x f_1 = \frac{1}{\delta_V}\mathcal{Q}_1(f_1) - \frac{1}{\delta_E}\mathcal{E}(f_0, f_1), \tag{II.11}$$

where now all the parameters ν_0, ν_1, d_0, d_1, τ_0, τ_1 are of order unity. Thus, we can investigate the two asymptotic regimes:

1. $\delta_V \ll \delta_E \ll 1$, when interactions among particles of the same group are more frequent than the exchanges between the two groups;
2. $\delta_E \ll \delta_V \ll 1$, when the exchanges between the groups are more frequent than interactions among particles of the same group.

Hydrodynamic rescaling. To study these two asymptotic regimes, let us perform a hydrodynamic rescaling by introducing new variables: $\tilde{x} = \varepsilon x$, $\tilde{t} = \varepsilon t$, with $\varepsilon \ll 1$. In the new variables, the distribution function $f^\varepsilon(\tilde{x}, \omega, \tilde{t}) = f(x, \omega, t)$ satisfies the system (dropping the tildes):

$$\varepsilon(\partial_t f_0^\varepsilon) = \frac{1}{\delta_V^\varepsilon}\mathcal{Q}_0^\varepsilon(f_0^\varepsilon) + \frac{1}{\delta_E^\varepsilon}\mathcal{E}^\varepsilon(f_0^\varepsilon, f_1^\varepsilon), \tag{II.12}$$

$$\varepsilon(\partial_t f_1^\varepsilon + \omega \cdot \nabla_x f_1^\varepsilon) = \frac{1}{\delta_V^\varepsilon}\mathcal{Q}_1^\varepsilon(f_1^\varepsilon) - \frac{1}{\delta_E^\varepsilon}\mathcal{E}^\varepsilon(f_0^\varepsilon, f_1^\varepsilon), \tag{II.13}$$

$$\mathcal{Q}_0^\varepsilon(f_0^\varepsilon) = -\nabla_\omega \cdot (\nu_0 \mathcal{F}^\varepsilon[f_0^\varepsilon] f_0^\varepsilon) + d_0 \Delta_\omega f_0^\varepsilon, \tag{II.14}$$

$$\mathcal{Q}_1^\varepsilon(f_1^\varepsilon) = -\nabla_\omega \cdot (\nu_1 \mathcal{F}^\varepsilon[f_1^\varepsilon] f_1^\varepsilon) + d_1 \Delta_\omega f_1^\varepsilon, \tag{II.15}$$

$$\mathcal{E}^\varepsilon(f_0^\varepsilon, f_1^\varepsilon) = -\frac{1}{\tau_0}\mathcal{G}^\varepsilon[f_1^\varepsilon] f_0^\varepsilon + \frac{1}{\tau_1}\mathcal{G}^\varepsilon[f_0^\varepsilon] f_1^\varepsilon, \tag{II.16}$$

where $\mathcal{F}^\varepsilon$ is the rescaled interaction forces given by:

$$\mathcal{F}^\varepsilon[f^\varepsilon](x, \omega, t) = (\mathrm{Id} - \omega \otimes \omega)\bar{\omega}^\varepsilon(x, \omega, t),$$
$$\bar{\omega}^\varepsilon(x, \omega, t) = \frac{J^\varepsilon(x, t)}{|J^\varepsilon(x, t)|}, \quad J^\varepsilon(x, t) = \frac{1}{\varepsilon^2}\int_{y \in \mathbb{R}^2, v \in \mathbb{S}^1} K\left(\left|\frac{x - y}{\varepsilon}\right|\right) v f^\varepsilon(y, v, t) dy dv,$$

and $\mathcal{G}^\varepsilon$ is the rescaled coupling coefficient:

$$\mathcal{G}^\varepsilon[f^\varepsilon](x, \omega, t) = 1 + \frac{\alpha}{\varepsilon^2}\int_{y \in \mathbb{R}^2, v \in \mathbb{S}^1} K\left(\left|\frac{x - y}{\varepsilon}\right|\right) \frac{(1 + v \cdot \omega)}{2} f^\varepsilon(y, v, t) dy dv. \tag{II.17}$$

Note that in these macroscopic variables, the effective interaction frequencies are $1/(\varepsilon\delta_V^\varepsilon)$ and $1/(\varepsilon\delta_E^\varepsilon)$.

The two asymptotics can be studied by the following assumptions: for the case $\delta_V \ll \delta_E \ll 1$, we can first consider the following asymptotics $\varepsilon\delta_V^\varepsilon = \varepsilon\delta_V = O(\varepsilon)$ and $\varepsilon\delta_E^\varepsilon = \delta_E = O(1)$ and then take the limit $\delta_E \to 0$. The first step consists in matching the scale of the macroscopic variables with the scale of the exchange interactions, while the second

step consists in considering that the time scales of these exchange interactions are also low with respect to the scale of the observation. Similarly, for the case $\delta_E \ll \delta_V \ll 1$, we can consider the asymptotics $\varepsilon\delta_E^\varepsilon = \varepsilon\delta_E = O(\varepsilon)$ and $\varepsilon\delta_V^\varepsilon = \delta_V = O(1)$ and then consider the limit $\delta_V \to 1$.

Before investigating these asymptotics, let us rewrite our model in a more convenient form.

Expansion with respect to ε. Let us introduce the macroscopic quantities related to a distribution function $f(x,\omega,t)$. The density $\rho[f](x,t)$, the momentum $j[f](x,t)$ and the mean direction $\Omega[f](x,t)$ are given by:

$$\rho[f](x,t) = \int_{v\in\mathbb{S}^1} f(x,v,t)dv, \quad j[f](x,t) = \int_{v\in\mathbb{S}^1} vf(x,v,t)dv, \quad \Omega[f](x,t) = \frac{j[f](x,t)}{|j[f](x,t)|}.$$

The following lemma provides an expansion of $\bar{\omega}^\varepsilon$ and g^ε with respect to ε.

Lemma 1. *We have the expansions*

$$\bar{\omega}^\varepsilon(x,\omega,t) = \Omega[f^\varepsilon](x,t) + O(\varepsilon^2), \tag{II.18}$$

$$\mathcal{G}^\varepsilon[f^\varepsilon] = 1 + \alpha\frac{\rho[f^\varepsilon] + \omega\cdot j[f^\varepsilon]}{2} + O(\varepsilon^2), \tag{II.19}$$

where $\rho[f^\varepsilon]$, $j[f^\varepsilon]$, $\Omega[f^\varepsilon]$ are the density, the momentum and the mean direction related to f^ε.

The proof of Lemma 1 is omitted. Lemma 1 enables to write system (II.12)-(II.13) as following:

$$\partial_t f_0^\varepsilon = \frac{1}{\varepsilon\delta_V^\varepsilon}Q_0(f_0^\varepsilon) + \frac{1}{\varepsilon\delta_E^\varepsilon}E(f_0^\varepsilon, f_1^\varepsilon) + O\left(\frac{\varepsilon^2}{\varepsilon\delta_V^\varepsilon}, \frac{\varepsilon^2}{\varepsilon\delta_E^\varepsilon}\right), \tag{II.20}$$

$$\partial_t f_1^\varepsilon + \omega\cdot\nabla_x f_1^\varepsilon = \frac{1}{\varepsilon\delta_V^\varepsilon}Q_1(f_1^\varepsilon) - \frac{1}{\varepsilon\delta_E^\varepsilon}E(f_0^\varepsilon, f_1^\varepsilon) + O\left(\frac{\varepsilon^2}{\varepsilon\delta_V^\varepsilon}, \frac{\varepsilon^2}{\varepsilon\delta_E^\varepsilon}\right), \tag{II.21}$$

where Q_0, Q_1 and E are here the operators of order $O(1)$ in the expansions of $\mathcal{Q}_0^\varepsilon$, $\mathcal{Q}_1^\varepsilon$ and $\mathcal{E}^\varepsilon$. They are given by:

$$Q_{0,1}(f_{0,1}^\varepsilon) = -\nabla_\omega\cdot(\nu_{0,1}F[f_{0,1}^\varepsilon]f_{0,1}^\varepsilon) + d_{0,1}\Delta_\omega f_{0,1}^\varepsilon, \tag{II.22}$$

$$E(f_0^\varepsilon, f_1^\varepsilon) = -\frac{1}{\tau_0}g[f_1^\varepsilon]f_0^\varepsilon + \frac{1}{\tau_1}g[f_0^\varepsilon]f_1^\varepsilon, \tag{II.23}$$

where the expressions of F and g result from lemma 1:

$$F[f](x,\omega,t) = (\mathrm{Id} - \omega\otimes\omega)\Omega[f](x,t), \tag{II.24}$$

$$g[f](x,\omega,t) = \left(1 + \alpha\frac{\rho[f](x,t) + \omega\cdot j[f](x,t)}{2}\right). \tag{II.25}$$

3 Case $\delta_V \ll \delta_E \ll 1$

The case $\delta_V \ll \delta_E \ll 1$ corresponds to a situation where the interaction frequency between particles with the same speed is higher than between particles with different speeds: the predominant dynamics are the Vicsek interactions inside each populations of particles. As mentionned above, we investigate this asymptotic regime in two steps:

1. consider that the macroscopic scales of interactions are satisfying $\varepsilon\delta_V^\varepsilon = \varepsilon$ and $\varepsilon\delta_E^\varepsilon = \delta_E = O(1)$ and then take the limit $\varepsilon \to 0$;
2. take the limit $\delta_E \to 0$.

So, the starting point is the following system:

$$\partial_t f_0^\varepsilon = \frac{1}{\varepsilon} Q_0(f_0^\varepsilon) + \frac{1}{\delta_E} E(f_0^\varepsilon, f_1^\varepsilon) + O(\varepsilon), \tag{II.26}$$

$$\partial_t f_1^\varepsilon + \omega \cdot \nabla_x f_1^\varepsilon = \frac{1}{\varepsilon} Q_1(f_1^\varepsilon) - \frac{1}{\delta_E} E(f_0^\varepsilon, f_1^\varepsilon) + O(\varepsilon). \tag{II.27}$$

3.1 Limit $\varepsilon\delta_V^\varepsilon = \varepsilon \to 0$

Equilibria. We suppose that $\delta_E = O(1)$ and we want to take the limit $\varepsilon \to 0$ in (II.26)-(II.27). Therefore, assuming that f_0^ε and f_1^ε converge to limits denoted by f_0 and f_1 as $\varepsilon \to 0$, we have $Q_0(f_0) = 0$ and $Q_1(f_1) = 0$. According to [2], the kernels of Q_0 and Q_1 are spanned by density and mean direction. The distributions f_0 and f_1 are more precisely given by:

$$f_0(x,\omega,t) = \rho_0(x,t) M_{\Omega_0(x,t)}(\omega), \quad f_1(x,\omega,t) = \rho_1(x,t) M_{\Omega_1(x,t)}(\omega),$$

where M_Ω is the so-called von Mises distribution:

$$M_{\Omega_{0,1}}(\omega) = C_{0,1} \exp\left(\frac{\omega \cdot \Omega_{0,1}}{\lambda_{0,1}}\right), \quad \lambda_{0,1} = \frac{d_{0,1}}{\nu_{0,1}},$$

where $C_{0,1}$ are normalization constants such that $\int_{\omega\in\mathbb{S}^1} M_{\Omega_{0,1}}(\omega) d\omega = 1$. In all the sequel, subscripts $_{0,1}$ mean that the left subscript $_0$ and the right subscript $_1$ are possible. The density, the momentum and the mean direction of these distributions are given by:

$$\rho[f_{0,1}](x,t) = \rho_{0,1}(x,t), \quad j[f_{0,1}](x,t) = c_{0,1}\rho_{0,1}(x,t)\Omega_{0,1}(x,t), \quad \Omega[f_{0,1}](x,t) = \Omega_{0,1}(x,t),$$

where $c_{0,1}$ are constants defined as $c_{0,1} = \langle \cos\theta \rangle_{M_{0,1}}$. For any function $s(\cos\theta)$, the brackets $\langle s(\cos\theta)\rangle_{M_{0,1}}$ will denote the average of s with respect to $M_{\Omega_{0,1}}$:

$$\langle s(\cos\theta)\rangle_{M_{0,1}} = \int_{\theta=0}^{2\pi} s(\cos\theta) C_{0,1} e^{\frac{\cos\theta}{\lambda_{0,1}}} d\theta.$$

$M_{\Omega_{0,1}}$ will be denoted $M_{0,1}$ in expressions expressed in polar coordinate with θ denoting the angle to $\Omega_{0,1}$.

Collisionnal invariants. We would like now to obtain the dynamics of the macroscopic quantities, which span the set of local equilibria. With this aim, the usual method is to integrate equations (II.26)-(II.27) against collisionnal invariants, which are quantities $m(\omega)$ belonging to the orthogonal of the image of $Q_{0,1}$. A condition to recover the dynamics of the equilibria is that the dimension of the vector space of collisionnal invariants equals the dimension of the vector space of local equilibria, which here is 2. It is not the case in the Vicsek dynamics since the only collisional invariant is mass: $m(\omega) = 1$. In [2], this difficulty is overcomed by considering generalized collisional invariants, that is collisional invariants ψ valid only for the subset of distribution function with prescribed mean direction:

$$\int_{\omega\in\mathbb{S}^1} Q_{0,1}(f)\psi = 0, \quad \forall f \text{ such that } \Omega[f] = \Omega.$$

In dimension 2, it has been shown in [2] that the vector space of the generalized collisional invariants for Q_0 (resp. Q_1) is spanned by m and a function ψ_0 (resp. ψ_1), which is the unique solution with zero average of the ellitic equation:

$$\partial_\theta\left(e^{\cos\theta/\lambda_0}\partial_\theta\psi_0\right) = \sin\theta e^{\cos\theta/\lambda_0}, \quad \left(\text{resp. } \partial_\theta\left(e^{\cos\theta/\lambda_1}\partial_\theta\psi_1\right) = \sin\theta e^{\cos\theta/\lambda_1}\right), \tag{II.28}$$

where we identified the functions on $\mathbb{S}^1$ and the 2π-periodic functions of $\mathbb{R}$, using polar coordinates in the basis $(\Omega, \Omega^\perp)$. Their explicit expressions are:

$$\psi_{0,1}(\theta) = \lambda_{0,1}\left(\pi\frac{\int_0^\theta e^{-\cos\varphi/\lambda_{0,1}}d\varphi}{\int_0^\pi e^{-\cos\varphi/\lambda_{0,1}}d\varphi} - \theta\right).$$

We define $h_{0,1}(\cos\theta) = \psi_{0,1}(\theta)/\sin\theta$ and so we have: $\psi_{0,1}(\omega) = h_{0,1}(\omega\cdot\Omega[f_{0,1}])\left(\Omega[f_{0,1}]^\perp\cdot\omega\right)$.

Mass equation. We integrate equations (II.26)-(II.27) (multiplied by the collisional invariant $m_1(\omega) = 1$) with respect to ω. We obtain:

$$\partial_t\rho_0^\varepsilon = \frac{1}{\delta_E}\int_{\omega\in\mathbb{S}^1} E(f_0^\varepsilon, f_1^\varepsilon)d\omega + O(\varepsilon), \tag{II.29}$$

$$\partial_t\rho_1^\varepsilon + \nabla_x\cdot j_1^\varepsilon = -\frac{1}{\delta_E}\int_{\omega\in\mathbb{S}^1} E(f_0^\varepsilon, f_1^\varepsilon)d\omega + O(\varepsilon). \tag{II.30}$$

So, in the limit $\varepsilon\to 0$, system (II.29)-(II.30) results in the following mass equations:

$$\partial_t\rho_0 = \frac{1}{\delta_E}R, \tag{II.31}$$

$$\partial_t\rho_1 + \nabla_x\cdot(c_1\rho_1\Omega_1) = -\frac{1}{\delta_E}R, \tag{II.32}$$

where the macroscopic exchange term R is a function of the macroscopic quantities:

$$R = \int_{\omega\in\mathbb{S}^1} E(\rho_0 M_{\Omega_0}, \rho_1 M_{\Omega_1})d\omega = \frac{\rho_1}{\tau_1} - \frac{\rho_0}{\tau_0} + \alpha\left(\frac{1}{\tau_1} - \frac{1}{\tau_0}\right)\rho_0\rho_1\Phi, \tag{II.33}$$

and Φ is the macroscopic alignement of the two populations:

$$\Phi = \left(\frac{1 + c_0c_1(\Omega_0\cdot\Omega_1)}{2}\right).$$

Momentum equation. We multiply (II.26)-(II.27) by the generalized collisional invariants $\psi_{0,1}(\omega)$ and we integrate with respect to ω. After some computations reported in appendix A (as done in [2]), we obtain in the limit $\varepsilon \to 0$:

$$\lambda_0^{-1}\langle(\sin\theta)^2 h_0\rangle_{M_{\Omega_0}}\rho_0\partial_t\Omega_0 = (\mathrm{Id}-\Omega_0\otimes\Omega_0)\left[-\langle(\sin\theta)^2 h_0\rangle_{M_{\Omega_0}}\nabla_x\rho_0 + \frac{1}{\delta_E}S_0\right], \quad \text{(II.34)}$$

$$\begin{aligned}\lambda_1^{-1}\langle(\sin\theta)^2 h_1\rangle_{M_{\Omega_1}}\rho_1\partial_t\Omega_1 + \lambda_1^{-1}\langle(\sin\theta)^2\cos\theta h_1\rangle_{M_{\Omega_1}}\rho_1(\Omega_1\cdot\nabla_x)\Omega_1 \\ = (\mathrm{Id}-\Omega_1\otimes\Omega_1)\left[-\langle(\sin\theta)^2 h_1\rangle_{M_{\Omega_1}}\nabla_x\rho_1 + \frac{1}{\delta_E}S_1\right],\end{aligned} \quad \text{(II.35)}$$

where S_0 and S_1 are the terms coming from the exchange operator:

$$S_0 = \int_{\omega\in\mathbb{S}^1}\left(-\frac{1}{\tau_0}g_1(\rho_1,\Omega_1)\rho_0 M_{\Omega_0} + \frac{1}{\tau_1}g_0(\rho_0,\Omega_0)\rho_1 M_{\Omega_1}\right)h_0(\omega\cdot\Omega_0)\omega d\omega, \quad \text{(II.36)}$$

$$S_1 = \int_{\omega\in\mathbb{S}^1}\left(\frac{1}{\tau_0}g_1(\rho_1,\Omega_1)\rho_0 M_{\Omega_0} - \frac{1}{\tau_1}g_0(\rho_0,\Omega_0)\rho_1 M_{\Omega_1}\right)h_1(\omega\cdot\Omega_1)\omega d\omega. \quad \text{(II.37)}$$

where g_0 and g_1 are expressed as function of the macroscopic quantities

$$g_{0,1}(\rho_{0,1},\Omega_{0,1}) = g[\rho_{0,1}M_{\Omega_{0,1}}] = 1 + \alpha\rho_{0,1}\frac{(1+c_{0,1}(\omega\cdot\Omega_{0,1}))}{2}.$$

Unlike the exchange terms in the mass equations, S_0 and S_1 are not opposite because of the presence of h_0 and h_1. However, they are always symetric in subscripts: when changing subscripts 0 to 1 and 1 to 0 in S_0, we obtain S_1.

Macroscopic system. We now collect the whole system in the following proposition:

Proposition 2. *The macroscopic quantities satisfy the equations:*

$$\partial_t\rho_0 = \frac{1}{\delta_E}R, \quad \text{(II.38)}$$

$$\partial_t\rho_1 + \nabla_x\cdot(c_1\rho_1\Omega_1) = -\frac{1}{\delta_E}R, \quad \text{(II.39)}$$

$$\rho_0\partial_t\Omega_0 = \lambda_0(Id-\Omega_0\otimes\Omega_0)\left[-\nabla_x\rho_0 + \frac{\beta_0^{-1}}{\delta_E}S_0\right], \quad \text{(II.40)}$$

$$\rho_1\partial_t\Omega_1 + \gamma_1\rho_1(\Omega_1\cdot\nabla_x)\Omega_1 = \lambda_1(Id-\Omega_1\otimes\Omega_1)\left[-\nabla_x\rho_1 + \frac{\beta_1^{-1}}{\delta_E}S_1\right], \quad \text{(II.41)}$$

where for the sake of clarity, we have introduce new parameters:

$$\gamma_1 = \frac{\langle(\sin\theta)^2\cos\theta h_1\rangle_{M_{\Omega_1}}}{\langle(\sin\theta)^2 h_1\rangle_{M_{\Omega_1}}}, \ \beta_{0,1} = \langle(\sin\theta)^2 h_{0,1}\rangle_{M_{\Omega_{0,1}}}.$$

The exchange term R in the mass equations is given in eq. (II.33). The exchange terms S_0 and S_1, in the direction equations, are given by equations (II.36)-(II.37).

3.2 Limit $\delta_E \to 0$: equilibria

Equilibria (for density). As δ_E goes to 0 in mass equations (II.38)-(II.39), we obtain:

$$R = 0,$$

which provides the balance equation for the densities

$$\rho_0 = f(\rho_1), \quad f(\rho_1) = \rho_1 \left(\frac{\tau_1}{\tau_0} + \alpha \left(\frac{\tau_1}{\tau_0} - 1 \right) \Phi \rho_1 \right)^{-1}. \tag{II.42}$$

Note that in the general case, this relation is not explicit since Φ depends on the directions Ω_0 and Ω_1, whose evolutions are non-linearly related to the densities via mass and momentum equations. The function f is increasing : it is represented in Fig. II.1. The following proposition gives the domain where both ρ_0 and ρ_1 are non negative.

Proposition 3. *(Conditions for positivity) If $\alpha > 0$ and supposing that ρ_0 and ρ_1 are non-negative, then the following results hold :*

1. *If $\tau_1/\tau_0 > 1$, then the density ρ_0 is bounded:*

$$\rho_0 \leqslant \frac{1}{\alpha\Phi\,(\tau_1/\tau_0 - 1)} \leqslant \frac{1}{\alpha\Phi_{\min}\,(\tau_1/\tau_0 - 1)},$$

   where $\Phi_{\min} = (1 - c_0 c_1)/2$.

2. *If $\tau_1/\tau_0 < 1$, then the density ρ_1 is bounded:*

$$\rho_1 \leqslant \frac{1}{\alpha\Phi\,(\tau_0/\tau_1 - 1)} \leqslant \frac{1}{\alpha\Phi_{\min}\,(\tau_0/\tau_1 - 1)}.$$

3. *If $\tau_1 = \tau_0$, then ρ_0 and ρ_1 are proportional. There is no restriction.*

If $\alpha = 0$, then the densities ρ_0 and ρ_1 are proportional. There is no restriction.

The proof of this proposition is omitted. This proposition shows that the dependency of the exchange rates on the local alignement implies that the population with the higher interaction frequency is bounded in time.

Finally, the density balance (II.42) can be equivalently expressed in terms of the total density $\rho = \rho_0 + \rho_1$:

$$\rho = k(\rho_1), \quad k(x) = x + f(x). \tag{II.43}$$

The function k is increasing.

Equilibria (for mean direction). Secondly, as δ_E goes to 0 in mean direction equations (II.40)-(II.41), we obtain:

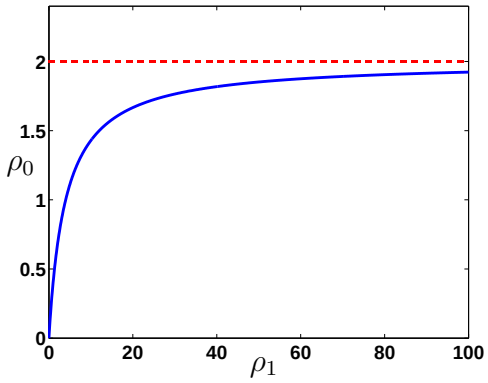
$$(\mathrm{Id} - \Omega_0 \otimes \Omega_0) S_0 = 0, \quad (\mathrm{Id} - \Omega_1 \otimes \Omega_1) S_1 = 0. \tag{II.44}$$

It can be easily checked from (II.36)-(II.37) that the most simple solutions are $\Omega_0 = \Omega_1$ or $\Omega_0 = -\Omega_1$, i.e. when the two populations are in the same or in the opposite direction. The following proposition states that they are the only two possible solutions.

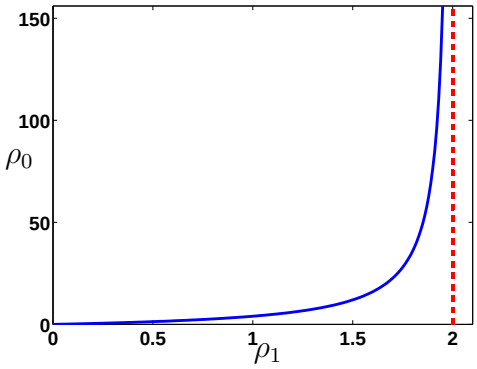
Proposition 4. *The only solutions to equations (II.44) are given by:*

$$\Omega_0 = \Omega_1 \quad \textit{and} \quad \Omega_0 = -\Omega_1.$$

The proof of this proposition can be found in appendix B. The question of the stability of the two direction balances will be addressed in future work.



(a) $\tau_1/\tau_0 = 2 > 1$



(b) $\tau_1/\tau_0 = 0.5 < 1$

Figure II.1: The relation $\rho_0 = f(\rho_1)$ with respect to the ratio τ_1/τ_0, with fixed $\phi = 0.5$ and $\alpha = 1$.

3.3 Limit $\delta_E \to 0$: equilibrium equations

In this section, δ will refer to the time scale δ_E of exchange between the two populations. We make the following expansions with respect to δ:

$$\rho_0^\delta = \rho_0 + \delta\tilde{\rho}_0 + O(\delta^2), \qquad \rho_1^\delta = \rho_1 + \delta\tilde{\rho}_1 + O(\delta^2),$$
$$\Omega_0^\delta = \Omega_0 + \delta\tilde{\Omega}_0 + O(\delta^2), \qquad \Omega_1^\delta = \Omega_1 + \delta\tilde{\Omega}_1 + O(\delta^2).$$

Let us note that Ω_0^δ and Ω_0 are of norm 1 and therefore $\tilde{\Omega}_0$ is orthogonal to Ω_0 (resp. $\tilde{\Omega}_1$ is orthogonal to Ω_1). Thus, we have the following expansions of the von-Mises distributions:

$$M_{\Omega_0^\delta}(\omega) = M_{\Omega_0}(\omega)(1+\delta\lambda_0^{-1}(\omega\cdot\tilde{\Omega}_0)+O(\delta^2)), \quad M_{\Omega_1^\delta}(\omega) = M_{\Omega_1}(\omega)(1+\delta\lambda_1^{-1}(\omega\cdot\tilde{\Omega}_1)+O(\delta^2)).$$

The following proposition states the expansion of the exchange terms R, S_0 and S_1 with respect to δ.

Lemma 5. ***1.*** *We have the following expansion:*

$$R^\delta = R + \delta\tilde{R} + O(\delta^2),$$

with

$$\tilde{R} = \frac{\tilde{\rho}_1}{\tau_1} - \frac{\tilde{\rho}_0}{\tau_0} + \alpha\left(\frac{1}{\tau_1} - \frac{1}{\tau_0}\right)(\tilde{\rho}_0\rho_1 + \rho_0\tilde{\rho}_1)\left(\frac{1 + c_0c_1\Omega_0\cdot\Omega_1}{2}\right)$$
$$+ \alpha\left(\frac{1}{\tau_1} - \frac{1}{\tau_0}\right)\rho_0\rho_1\frac{c_0c_1}{2}\left(\tilde{\Omega}_0\cdot\Omega_1 + \Omega_0\cdot\tilde{\Omega}_1\right).$$

2. *We have the following expansions:*

$$S_0^\delta = S_0 + \delta\tilde{S}_0 + 0(\delta^2), \tag{II.45}$$
$$(Id - \Omega_0^\delta\otimes\Omega_0^\delta)S_0^\delta = (Id - \Omega_0\otimes\Omega_0)S_0 + \delta(Id - \Omega_0\otimes\Omega_0)\tilde{S}_0$$
$$-\delta(\tilde{\Omega}_0\cdot S_0)\Omega_0 - \delta(\Omega_0\cdot S_0)\tilde{\Omega}_0 + O(\delta^2), \tag{II.46}$$

with

$$\begin{aligned}
\tilde{S}_0 &= \frac{1}{\tau_1}X_0 - \frac{1}{\tau_0}Y_0,\\
X_0 &= \int_{\omega\in\mathbb{S}^1} \tilde{g}_0(\rho_0,\Omega_0)(\tilde{\rho}_0,\tilde{\Omega}_0)\rho_1 M_{\Omega_1} h_0(\omega\cdot\Omega_0)\omega d\omega + \int_{\omega\in\mathbb{S}^1} g_0(\rho_0,\Omega_0)(\widetilde{\rho_1 M_{\Omega_1}})h_0(\omega\cdot\Omega_0)\omega d\omega\\
&\quad + \int_{\omega\in\mathbb{S}^1} g_0(\rho_0,\Omega_0)\rho_1 M_{\Omega_1} h_0'(\omega\cdot\Omega_0)(\omega\otimes\omega)d\omega\ \tilde{\Omega}_0,\\
Y_0 &= \int_{\omega\in\mathbb{S}^1} \tilde{g}_1(\rho_1,\Omega_1)(\tilde{\rho}_1,\tilde{\Omega}_1)\rho_0 M_{\Omega_0} h_0(\omega\cdot\Omega_0)\omega d\omega + \int_{\omega\in\mathbb{S}^1} g_1(\rho_1,\Omega_1)(\widetilde{\rho_0 M_{\Omega_0}})h_0(\omega\cdot\Omega_0)\omega d\omega\\
&\quad + \int_{\omega\in\mathbb{S}^1} g_1(\rho_1,\Omega_1)\rho_0 M_{\Omega_0} h_0'(\omega\cdot\Omega_0)(\omega\otimes\omega)d\omega\ \tilde{\Omega}_0,
\end{aligned}$$

and

$$\begin{aligned}
&\tilde{g}_{0,1}(\rho_{0,1},\Omega_{0,1})(\tilde{\rho}_{0,1},\tilde{\Omega}_{0,1}) = \alpha\left(\frac{\tilde{\rho}_{0,1} + c_{0,1}(\tilde{\rho}_{0,1}\Omega_{0,1} + \rho_{0,1}\tilde{\Omega}_{0,1})\cdot\omega}{2}\right),\\
&(\widetilde{\rho_{0,1}M_{\Omega_{0,1}}}) = \left(\tilde{\rho}_{0,1} + \rho_{0,1}\lambda_{0,1}^{-1}(\omega\cdot\tilde{\Omega}_{0,1})\right)M_{\Omega_{0,1}}.
\end{aligned}$$

By simply changing subscripts 0 *into* 1 *and* 1 *into* 0*, we obtain the expansion of* S_1^δ*.*

The proof of this lemma is easy and is omitted.

We now investigate the equilibrium equations by inserting the previous expansions into equations (II.38)-(II.39)-(II.40)-(II.41).

- At the leading order $O(\delta^{-1})$, we obtain:

$$R = 0, \quad (\mathrm{Id} - \Omega_0\otimes\Omega_0)S_0 = 0, \quad (\mathrm{Id} - \Omega_1\otimes\Omega_1)S_1 = 0.$$

Consequently, according to the previous section, the following relations hold:

$$\rho = k(\rho_1), \quad \Omega_0 = \Omega_1 \text{ or } \Omega_0 = -\Omega_1,$$

where ρ is the total density and the function k is defined by (II.43) and as the following form:

$$k(x) = x + x\left(\frac{\tau_1}{\tau_0} + \alpha\left(\frac{\tau_1}{\tau_0} - 1\right)\Phi x\right)^{-1}. \tag{II.47}$$

The relations on the directions are provided by proposition 4 and imply that ϕ is a constant, given by:

$$\Phi = \frac{(1 + c_0c_1)}{2}, \quad \text{if } \Omega_0 = \Omega_1,$$

or

$$\Phi = \frac{(1 - c_0c_1)}{2}, \quad \text{if } \Omega_0 = -\Omega_1.$$

- At order $O(1)$, we obtain:

$$\partial_t\rho_0 = \tilde{R}, \tag{II.48}$$

$$\partial_t\rho_1 + \nabla_x\cdot(c_1\rho_1\Omega_1) = -\tilde{R}, \tag{II.49}$$

$$\begin{aligned}
&\rho_0\partial_t\Omega_0 + \lambda_0(\mathrm{Id} - \Omega_0\otimes\Omega_0)\nabla_x\rho_0 =\\
&\qquad \lambda_0\beta_0^{-1}(\mathrm{Id} - \Omega_0\otimes\Omega_0)\tilde{S}_0 - \lambda_0\beta_0^{-1}(\Omega_0\cdot S_0)\tilde{\Omega}_0,
\end{aligned} \tag{II.50}$$

$$\begin{aligned}
&\rho_1\partial_t\Omega_1 + \gamma_1\rho_1(\Omega_1\cdot\nabla_x)\Omega_1 + \lambda_1(\mathrm{Id} - \Omega_1\otimes\Omega_1)\nabla_x\rho_1 =\\
&\qquad \lambda_1\beta_1^{-1}(\mathrm{Id} - \Omega_1\otimes\Omega_1)\tilde{S}_1 - \lambda_1\beta_1^{-1}(\Omega_1\cdot S_1)\tilde{\Omega}_1,
\end{aligned} \tag{II.51}$$

where $\tilde{R}$, $\tilde{S}_0$, $\tilde{S}_1$ are the first order correction of the exchange term R, S_0 and S_1 and their expressions are given in lemma 5. They should be expressed around equilibria. Note that the terms $(\tilde{\Omega}_0 \cdot S_0)\Omega_0$ (resp. $(\tilde{\Omega}_1 \cdot S_1)\Omega_1$) in (II.46) do not appear in equation (II.50) (resp. (II.51)) since S_0 is parallel to Ω_0 (resp. S_1 is parallel to Ω_1).

System (II.48)-(II.49)-(II.50)-(II.51) is not closed: the dynamics of the first order correction terms (depending on $\tilde{\rho}_0$, $\tilde{\rho}_1$, $\tilde{\Omega}_0$, $\tilde{\Omega}_1$) should be provided by the equations at order $O(\delta)$. However, by expressing compatibility conditions, we are able to close the system. We consider the case $\Omega_0 = \Omega_1$.

Actually, adding the two density equations, we get the following closed equation:

$$\partial_t(\rho_0 + \rho_1) + \nabla_x \cdot (c_1\rho_1\Omega_1) = 0,$$

where the exchange terms have been concelled.

For equations (II.50)-(II.51), the operator at the right-hand side of equations (II.50)-(II.51) is a linear operator acting on $\tilde{\Omega}_0, \tilde{\Omega}_1 \in \text{vect}(\Omega_0^\perp)$, where $\text{vect}(\Omega_0^\perp)$ denotes the line spanned by $\Omega_0^\perp$. Its Null-Space is not reduced to $\{0\}$ since it is not the case for the original operator. Thus a closed equation can be obtained just by expressing that the left-hand side of equations (II.50)-(II.51) have to belong to the one-dimensional image of the linearized operator. The following proposition gives the explicit expression of this linear operator.

Proposition 6. *(Case $\Omega_0 = \Omega_1$) We have:*

$$\begin{aligned}
&\forall \rho_0, \rho_1, \tilde{\rho}_0, \tilde{\rho}_1 \in \mathbb{R}^+,\ \forall \Omega_0 = \Omega_1 \in \mathbb{S}^1,\ \forall \tilde{\Omega}_0, \tilde{\Omega}_1 \in \textit{vect}(\Omega_0^\perp),\\
&\qquad \lambda_0\beta_0^{-1}(Id - \Omega_0 \otimes \Omega_0)\tilde{S}_0 - \lambda_0\beta_0^{-1}(\Omega_0 \cdot S_0)\tilde{\Omega}_0 = A_0(\tilde{\Omega}_1 - \tilde{\Omega}_0),\\
&\qquad \lambda_1\beta_1^{-1}(Id - \Omega_1 \otimes \Omega_1)\tilde{S}_1 - \lambda_1\beta_1^{-1}(\Omega_1 \cdot S_1)\tilde{\Omega}_1 = A_1(\tilde{\Omega}_0 - \tilde{\Omega}_1),
\end{aligned}$$

where A_0 and A_1 are two functions denpending only on ρ_0 and ρ_1 and given by:

$$\begin{aligned}
A_0 = \lambda_0\beta_0^{-1}\Big(&\frac{1}{\tau_1}(1 + \alpha\frac{\rho_0}{2})\rho_1\lambda_1^{-1}\langle \sin^2\theta h_0\rangle_{M_1} + \frac{1}{\tau_1}\alpha\frac{c_0\rho_0}{2}\rho_1\lambda_1^{-1}\langle \sin^2\theta\cos\theta h_0\rangle_{M_1}\\
&-\frac{1}{\tau_0}\alpha\frac{c_1\rho_1}{2}\rho_0\langle \sin^2\theta h_0\rangle_{M_0}\Big),
\end{aligned}$$

and A_1 with the same expression by changing 1 to 0 and 0 to 1.

The proof of this proposition is developped in appendix C. It is now an easy matter to derive a closed system :

Proposition 7. *In the case where $\Omega_0 = \Omega_1$, system (II.48)-(II.49)-(II.50)-(II.51) yields the following closed system for the total density $\rho = \rho_0 + \rho_1$ and the common mean direction $\Omega = \Omega_0 = \Omega_1$:*

$$\partial_t\rho + \nabla_x \cdot (c_1 k^{-1}(\rho)\Omega) = 0, \tag{II.52}$$

$$\begin{aligned}
&(A_1\rho + (A_0 - A_1)k^{-1}(\rho))\ \partial_t\Omega + A_0\gamma_1 k^{-1}(\rho)\ (\Omega \cdot \nabla_x)\Omega =\\
&\qquad -(\lambda_0 A_1 + (\lambda_1 A_0 - \lambda_0 A_1)(k^{-1})'(\rho))(Id - \Omega \otimes \Omega)\ \nabla_x\rho,
\end{aligned} \tag{II.53}$$

where A_0 and A_1 are defined in proposition 6 and k is given by (II.47).

System (II.52)-(II.53) provides the dynamics of the two populations at equilibrium. The study of its mathematical properties will be the subject of future work.

Remark: The derivation of this closed system can be interpreted with a "collisional invariant" viewpoint. To obtain the density equation, we use that the exhange interactions preserve mass and thus the vector $(R, -R)$ as well as its linearized operator $(\tilde{R}, -\tilde{R})$ satifies:

$$\forall \rho_0, \rho_1 \in \mathbb{R}^+,\ \forall \Omega_0, \Omega_1 \in \mathbb{S}^1, \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} R \\ -R \end{pmatrix} = 0.$$

Since momentum is not conserved, such a relation is lacking for the momentum exchange operator $(\mathrm{Id} - \Omega_0 \otimes \Omega_0)S_0, (\mathrm{Id} - \Omega_1 \otimes \Omega_1)S_1$. However, the previous results show that the linearized operator around an equilibria have such a collisionnal invariant (equal to (A_0, A_1)) and that it is sufficient to conclude.

4 Case $\delta_E \ll \delta_Q \ll 1$

The case $\delta_E \ll \delta_V \ll 1$ corresponds to a situation where the frequency at which particles change their speeds is higher than the interaction frequency between particles with the same speed. As mentionned above, we investigate these asymptotics in two steps:

1. consider that macroscopic scales of interactions are satisfying $\varepsilon\delta_V^\varepsilon = \delta_V = O(1)$ and $\varepsilon\delta_E^\varepsilon = \varepsilon$ and then take the limit $\varepsilon \to 0$;
2. take the limit $\delta_V \to 0$.

So, the starting point is the following system:

$$\partial_t f_0^\varepsilon = \frac{1}{\delta_V} Q_0(f_0^\varepsilon) + \frac{1}{\varepsilon} E(f_0^\varepsilon, f_1^\varepsilon) + O(\varepsilon), \tag{II.54}$$

$$\partial_t f_1^\varepsilon + \omega \cdot \nabla_x f_1^\varepsilon = \frac{1}{\delta_V} Q_1(f_1^\varepsilon) - \frac{1}{\varepsilon} E(f_0^\varepsilon, f_1^\varepsilon) + O(\varepsilon). \tag{II.55}$$

The investigation of the macroscopic dynamics in this case is far from being complete. We will simply here raise some differences with the previous asymptotics.

4.1 Limit $\varepsilon\delta_E^\varepsilon = \varepsilon \to 0$

To take the limit $\varepsilon \to 0$ in (II.54)-(II.55), we make an Hibert expansion:

$$f_0^\varepsilon = f_0 + \varepsilon \tilde{f}_0 + O(\varepsilon^2), \quad f_1^\varepsilon = f_1 + \varepsilon \tilde{f}_1 + O(\varepsilon^2).$$

Since the functional E is a combination of a linear operator and a bilinear one, we easily get the following expansion:

$$\begin{aligned}
&E(f_0^\varepsilon, f_1^\varepsilon) = E(f_0, f_1) + \varepsilon dE(f_0, f_1)(\tilde{f}_0, \tilde{f}_1) + O(\varepsilon^2) \\
&dE(f_0, f_1)(\tilde{f}_0, \tilde{f}_1) = -\frac{1}{\tau_0}\left(\tilde{f}_0 g[f_1] + f_0 \tilde{g}[\tilde{f}_1]\right) + \frac{1}{\tau_1}\left(\tilde{f}_1 g[f_0] + f_1 \tilde{g}[\tilde{f}_0]\right), \\
&\tilde{g}[f] = \alpha \frac{\rho[f] + \omega \cdot j[f]}{2}.
\end{aligned}$$

Inserting these expansions into equations (II.12)-(II.17), we obtain at order $O(\varepsilon^{-1})$ the following equilibrium condition:

$$E(f_0, f_1) = 0,$$

that is

$$\frac{1}{\tau_0} g[f_1] f_0 = \frac{1}{\tau_1} g[f_0] f_1. \tag{II.56}$$

In the particular case $\alpha = 0$, this condition implies the equilibrium: $f_0 = \frac{\tau_0}{\tau_1} f_1$. In the case $\tau_0 = \tau_1$, the equality of the two distribution $f_0 = f_1$ is a particular solution. The question whether there exist equilibria in the case $\tau_0 \neq \tau_1$ and $\alpha \neq 0$ is open. At order $O(1)$, equations (II.12)-(II.17) lead to:

$$\partial_t f_0 = \frac{1}{\delta_V} Q_0(f_0) + dE(f_0, f_1)(\tilde{f}_0, \tilde{f}_1), \tag{II.57}$$

$$\partial_t f_1 + \omega \cdot \nabla_x f_1 = \frac{1}{\delta_V} Q_1(f_1) - dE(f_0, f_1)(\tilde{f}_0, \tilde{f}_1). \tag{II.58}$$

This system is not closed. To close it, we can sum the two equations:

$$\partial_t (f_1 + f_0) + \omega \cdot \nabla_x f_1 = \frac{1}{\delta_V} \left(Q_1(f_1) + Q_0(f_0) \right),$$

With the balance equation (II.56), it provides a closed equation for f_1 or f_0. In the case $\alpha = 0$, it leads to:

$$(1 + \frac{\tau_0}{\tau_1}) \partial_t f_1 + \omega \cdot \nabla_x f_1 = \frac{1}{\delta_V} Q(f_1), \tag{II.59}$$

where Q denotes the following Vicsek operator:

$$Q(f) = -\nabla_\omega \cdot ((\frac{\tau_0}{\tau_1} \nu_0 + \nu_1) F[f] f) + (\frac{\tau_0}{\tau_1} d_0 + d_1) \Delta_\omega f. \tag{II.60}$$

4.2 Limit $\delta_V \to 0$

The parameter δ will refer in this section to δ_V. We consider the expansion with respect to δ:

$$f_0^\delta = f_0 + O(\delta), \quad f_1^\delta = f_1 + O(\delta),$$

Taking the limit $\delta \to 0$ in (II.57)-(II.58) leads to:

$$Q_0(f_0) = 0, \quad Q_1(f_1) = 0.$$

The solutions of these latter equalities are von Mises velocity distributions:

$$f_0(x, \omega, t) = \rho_0 M_{\Omega_0}, \quad f_1(x, \omega, t) = \rho_1 M_{\Omega_1},$$
$$M_{\Omega_0} = C_0 \exp(\frac{\omega \cdot \Omega_0}{\lambda_0}), \quad M_{\Omega_1} = C_1 \exp(\frac{\omega \cdot \Omega_1}{\lambda_1}),$$

where C_0 and C_1 are renormalization constants. Integrating the balance (II.56) with respect to ω, we obtain the following mass balance:

$$\frac{\rho_0}{\tau_0} - \frac{\rho_1}{\tau_1} + \alpha \left(\frac{1}{\tau_0} - \frac{1}{\tau_1} \right) \rho_0 \rho_1 \left(\frac{1 + c_0 c_1 (\Omega_0 \cdot \Omega_1)}{2} \right) = 0, \tag{II.61}$$

with $c_0 = \langle \cos\theta \rangle_{M_{\Omega_0}}$ and $c_1 = \langle \cos\theta \rangle_{M_{\Omega_1}}$. It is the same as the one obtained in (II.42).

In the case $\alpha = 0$, the coupling of the balance equation (II.56) and the convergence to von Mises distributions need the following equalities to be satisfied:

$$\frac{\rho_0}{\tau_0} = \frac{\rho_1}{\tau_1}, \quad \Omega_0 = \Omega_1, \quad \lambda_0 = \lambda_1.$$

These conditions are more stringent than those obtained in section 3.2 and they proved that in the case $\lambda_0 \neq \lambda_1$, no equilibrium can be reached. Then, the integration of (II.59) against the generalized collisional invariants associated to Q leads to the following system (with the same computations already done in appendix A):

$$\left(1 + \frac{\tau_0}{\tau_1}\right) \partial_t \rho_1 + \nabla_x \cdot (c_1 \rho_1 \Omega_1) = 0, \tag{II.62}$$

$$\left(1 + \frac{\tau_0}{\tau_1}\right) \rho_1 \partial_t \Omega_1 + \gamma \rho_1 (\Omega_1 \cdot \nabla_x)\Omega_1 + \lambda_1 (\mathrm{Id} - \Omega_1 \otimes \Omega_1)\nabla_x \rho_1 = 0. \tag{II.63}$$

where γ is defined by:

$$\gamma = \frac{\langle \cos\theta (\sin\theta)^2 h \rangle_{M_{\Omega_1}}}{\langle (\sin\theta)^2 \rangle_{M_{\Omega_1}}},$$

where $h(\cos\theta) = \Psi(\theta)/\sin\theta$, where Ψ is the collisionnal invariant associated to Q, defined by (II.60). Note that the definition of h is here depending on the parameter λ given by:

$$\lambda = \frac{\tau_0 \nu_0 + \tau_1 \nu_1}{\tau_0 d_0 + \tau_1 d_1}.$$

and thus is different from $h_0 = h_1$. We can also check that system (II.62)-(II.63) is similar to the one obtained in the first asymptotic regime (system (II.52)-(II.53), when assuming that $\lambda_0 = \lambda_1$) except that the constant γ is replaced by γ_1.

The dynamics for cases $\alpha \neq 0$ is still under investigation.

5 Conclusion

In this chapter, we have studied the dynamics at large time and space scale of a two-phase model for particles with alignement interactions: one phase contains particles at rest with speed zero and one phase contains moving particles with speed one. Inside each phase, the particles follow the Vicsek dynamics. The exchange of particles between the two phases also depends on the local alignement of the two populations. We are interested in the dynamics of the system when the time scales of both the Vicsek interactions and the exchanges are small in comparison with the scale of observation. The first asymptotics we have investigated consist in considering that the interactions in each phase are more frequent than the exchanges between the two populations. The equilibria are found: the two population are aligned or in opposite direction. Moreover, the dynamics of the densities and the common direction are provided by cancelling the exchange terms. For the second asymptotic regime, the exchanges between the two populations are more frequent than the interactions in each phase. This leads to more stringent conditions to reach equilibria. Only the case where exchanges are independant of local alignement

has been successfully resolved and leads to a different macroscopic model from the one predicted by the first asymptotics.

The analysis is not totally complete. The stability of equilibria, for the first asymptotics, and their existence, for the second asymptotic, should be questionned. Numerical simulations will be carried out to valid this study.

A Appendix : Momentum equations (details)

For the sake of completeness, we present here the derivation of the evolution equation of Ω_1, given by (II.35). The derivation of equation (II.34) is similar.

We multiply system (II.21) by the generalized collisional invariant:

$$\psi_1^\varepsilon = h_1\left(\omega\cdot\Omega[f_1^\varepsilon]\right)\left(\Omega[f_1^\varepsilon]^\perp\cdot\omega\right)\Omega[f_1^\varepsilon]^\perp,$$

introduced in [2] and we integrate it with respect to ω. In the limit $\varepsilon\to 0$, we have $\Omega[f^\varepsilon]\to\Omega_1$ and $\psi_1^\varepsilon\to h_1\left(\omega\cdot\Omega_1\right)\left(\Omega_1^\perp\cdot\omega\right)\Omega_1^\perp$. Therefore, we obtain:

$$(\Omega_1^\perp\otimes\Omega_1^\perp)X=0,\quad X=\int_{\omega\in\mathbb{S}^1}\left[\partial_t(\rho_1 M_{\Omega_1})+\omega\cdot\nabla_x(\rho_1 M_{\Omega_1})+\frac{1}{\delta_E}E(\rho_0 M_{\Omega_0},\rho_1 M_{\Omega_1})\right]h_1\omega d\omega,$$

where h_1 and M_{Ω_1} are functions of $(\omega\cdot\Omega_1)$ and M_{Ω_0} depends only on $(\omega\cdot\Omega_0)$. The derivative of M_{Ω_1} with respect to Ω_1 acting on a tangent vector $d\Omega$ to the circle is given by:

$$\frac{\partial M_{\Omega_1}}{\partial\Omega_1}(d\Omega)=\lambda_1^{-1}(\omega\cdot\Omega_1)(\omega\cdot d\Omega)M_{\Omega_1}.$$

Thus, we have:

$$X = \int_{\omega\in\mathbb{S}^1}\left[\partial_t\rho_1+\omega\cdot\nabla_x\rho_1+\lambda_1^{-1}\rho_1\left(\omega\cdot\partial_t\Omega_1+(\omega\otimes\omega):\nabla_x\Omega_1\right)\right]M_{\Omega_1}h_1\omega d\omega-\frac{1}{\delta_E}S_1,$$

where S_1 is given by (II.37). The symbol ':' denotes the contracted product of two tensors (if $A=(A_{i,j})_{i,j=1,2}$ and $B=(B_{i,j})_{i,j=1,2}$ are two tensors then $A:B=\sum_{i,j=1,2}A_{i,j}B_{i,j}$) and $\nabla_x\Omega$ is the gradient tensor of the vector Ω: $(\nabla_x\Omega)_{i,j}=\partial_{x_i}\Omega_j$. The four first terms in this formula, denoted X_1 to X_4, are computed using polar coordinates (θ) related to the cartesian basis $(\Omega_1,\Omega_1^\perp)$: in this basis h_1 and M_{Ω_1} are even functions depending on $\cos\theta$.

By writing $\omega=(\omega\cdot\Omega_1)\Omega_1+(\omega\cdot\Omega_1^\perp)\Omega_1^\perp$, we decompose X_1 as follows:

$$\begin{aligned}X_1 &= \int_{\omega\in\mathbb{S}^1}\partial_t\rho_1 M_{\Omega_1}h_1\omega d\omega\\ &= \partial_t\rho_1\int_{\theta=-\pi}^{\pi}h_1 M_{\Omega_1}\cos\theta d\theta\ \Omega_1+\partial_t\rho_1\int_{\theta=-\pi}^{\pi}h_1 M_{\Omega_1}\sin\theta d\theta\ \Omega_1^\perp,\end{aligned}$$

The second term is zero (the integration of an odd function on a symetric interval) and thus we obtain:

$$(\Omega_1^\perp\otimes\Omega_1^\perp)X_1=0. \tag{II.64}$$

A similar computation shows that:

$$\begin{aligned} X_2 &= \int_{\omega\in\mathbb{S}^1} \left[(\omega\otimes\omega)\nabla_x\rho_1\right] M_{\Omega_1} h_1 d\omega \\ &= \int_{\theta=-\pi}^{\pi} h_1 M_{\Omega_1} \cos^2\theta d\theta\ (\Omega_1\otimes\Omega_1)\nabla_x\rho_1 \\ &\quad + \int_{\theta=-\pi}^{\pi} h_1 M_{\Omega_1} \cos\theta\sin\theta d\theta\ [(\Omega_1\otimes\Omega_1^\perp)+(\Omega_1^\perp\otimes\Omega_1)]\nabla_x\rho_1 \\ &\quad + \int_{\theta=-\pi}^{\pi} h_1 M_{\Omega_1} \sin^2\theta d\theta\ (\Omega_1^\perp\otimes\Omega_1^\perp)\nabla_x\rho_1. \end{aligned}$$

The first term is parallel to Ω_1 and the second term is zero. Therefore, we obtain:

$$(\Omega_1^\perp\otimes\Omega_1^\perp)X_2 = \langle\sin^2\theta h_1\rangle_{M_{\Omega_1}}(\Omega_1^\perp\otimes\Omega_1^\perp)\nabla_x\rho_1. \tag{II.65}$$

With similar computations, we obtain the following results for X_3:

$$\begin{aligned} X_3 &= \lambda_1^{-1}\rho_1 \int_{\omega\in\mathbb{S}^1} \left[(\omega\otimes\omega)\cdot\partial_t\Omega_1\right] M_{\Omega_1} h_1 d\omega, \\ (\Omega_1^\perp\otimes\Omega_1^\perp)X_3 &= \lambda_1^{-1}\rho_1\langle\sin^2\theta h_1\rangle_{M_{\Omega_1}}(\Omega_1^\perp\otimes\Omega_1^\perp)\partial_t\Omega_1. \end{aligned} \tag{II.66}$$

For X_4, we have:

$$\begin{aligned} X_4 &= \lambda_1^{-1}\rho \int_{\omega\in\mathbb{S}^1} \left[(\omega\otimes\omega):\nabla_x\Omega_1\right] h_1\omega d\omega \\ &= \lambda_1^{-1}\rho_1 \int_{\theta=-\pi}^{\pi} \Big[\cos^2\theta(\Omega_1\otimes\Omega_1) + \sin\theta\cos\theta(\Omega_1\otimes\Omega_1^\perp) \\ &\qquad + \sin\theta\cos\theta(\Omega_1^\perp\otimes\Omega_1) + \sin^2\theta(\Omega_1^\perp\otimes\Omega_1^\perp)\Big] : \nabla_x\Omega_1\ h_1 M_{\Omega_1}\cos\theta d\theta\ \Omega_1 \\ &\quad + \lambda_1^{-1}\rho_1 \int_{\theta=-\pi}^{\pi} \Big[\cos^2\theta(\Omega_1\otimes\Omega_1) + \sin\theta\cos\theta(\Omega_1\otimes\Omega_1^\perp) \\ &\qquad + \sin\theta\cos\theta(\Omega_1^\perp\otimes\Omega_1) + \sin^2\theta(\Omega_1^\perp\otimes\Omega_1^\perp)\Big] : \nabla_x\Omega_1\ h_1 M_{\Omega_1}\sin\theta d\theta\ \Omega^\perp. \end{aligned}$$

The odd function of θ vanishes and thus we are left with:

$$\begin{aligned} X_4 &= \lambda_1^{-1}\rho_1 \int_{\theta=-\pi}^{\pi} \Big[\cos^2\theta(\Omega_1\otimes\Omega_1) + \sin^2\theta(\Omega_1^\perp\otimes\Omega_1^\perp)\Big] : \nabla_x\Omega_1\ h_1 M_{\Omega_1}\cos\theta d\theta\ \Omega_1 \\ &\quad + \lambda_1^{-1}\rho_1 \int_{\theta=-\pi}^{\pi} \sin\theta\cos\theta\ \Big[(\Omega_1\otimes\Omega_1^\perp)+(\Omega_1^\perp\otimes\Omega_1)\Big] : \nabla_x\Omega\ h_1 M_{\Omega_1}\sin\theta d\theta\ \Omega_1^\perp. \end{aligned}$$

Besides, we have for all $\Omega\in\mathbb{S}^1$:

$$(\Omega\otimes\Omega^\perp):\nabla\Omega = \Omega_i\Omega_j^\perp\partial_i\Omega_j = ((\Omega\cdot\nabla)\Omega)\cdot\Omega^\perp,$$

and

$$(\Omega^\perp\otimes\Omega):\nabla\Omega = \Omega_i^\perp\Omega_j\partial_i\Omega_j = \frac{1}{2}\Omega^\perp\cdot\nabla|\Omega|^2 = 0.$$

So, we finally have

$$(\Omega_1^\perp\otimes\Omega_1^\perp)X_4 = \lambda_1^{-1}\rho\langle\sin^2\theta\cos\theta h_1\rangle_{M_{\Omega_1}}(\Omega_1^\perp\otimes\Omega_1^\perp)((\Omega_1\cdot\nabla)\Omega_1). \tag{II.67}$$

Equations (II.64)-(II.65)-(II.66)-(II.67) results in the momentum equations (II.35).

B Appendix: Proof of proposition 4 (Momenta balance)

Let us first explicit the expressions of S_0 and S_1 given by (II.36)-(II.37), when they are projected onto the lines spanned respectively by $\Omega_0^\perp$ and $\Omega_1^\perp$:

$$\begin{aligned}(\mathrm{Id}-\Omega_0\otimes\Omega_0)S_0 &= -(\mathrm{Id}-\Omega_0\otimes\Omega_0)(1+\alpha\frac{\rho_1}{2})\frac{\rho_0}{\tau_0}\int_{\omega\in\mathbb{S}^1} M_{\Omega_0}h_0(\omega\cdot\Omega_0)\omega d\omega\\ &\quad -(\mathrm{Id}-\Omega_0\otimes\Omega_0)\alpha\frac{c_1\rho_1}{2}\frac{\rho_0}{\tau_0}\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_0(\omega\cdot\Omega_0)d\omega\Omega_1\\ &\quad +(\mathrm{Id}-\Omega_0\otimes\Omega_0)(1+\alpha\frac{\rho_0}{2})\frac{\rho_1}{\tau_1}\int_{\omega\in\mathbb{S}^1} M_{\Omega_1}h_0(\omega\cdot\Omega_0)\omega d\omega\\ &\quad +(\mathrm{Id}-\Omega_0\otimes\Omega_0)\alpha\frac{c_0\rho_0}{2}\frac{\rho_1}{\tau_1}\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0(\omega\cdot\Omega_0)d\omega\Omega_0.\end{aligned}$$

The first term vanishes and the integral in the second term can be written as follows:

$$\begin{aligned}\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_0(\omega\cdot\Omega_0)d\omega &= \int_{\theta=0}^{2\pi}\cos^2\theta M_0(\cos\theta)h_0(\cos\theta)d\theta\ \Omega_0\otimes\Omega_0\\ &\quad +\int_{\theta=0}^{2\pi}\sin^2\theta M_0(\cos\theta)h_0(\cos\theta)d\theta\ (\mathrm{Id}-\Omega_0\otimes\Omega_0).\end{aligned}$$

We deduce a simplified expression for the projection of S_0:

$$\begin{aligned}(\mathrm{Id}-\Omega_0\otimes\Omega_0)S_0 &= -\alpha\frac{c_1\rho_1}{2}\frac{\rho_0}{\tau_0}\langle(\sin\theta)^2h_0\rangle_{M_{\Omega_0}}(\mathrm{Id}-\Omega_0\otimes\Omega_0)\Omega_1\\ &\quad +(1+\alpha\frac{\rho_0}{2})\frac{\rho_1}{\tau_1}\int_{\omega\in\mathbb{S}^1} M_{\Omega_1}h_0(\omega\cdot\Omega_0)(\omega\cdot\Omega_0^\perp)d\omega\Omega_0^\perp \qquad \text{(II.68)}\\ &\quad +\alpha\frac{c_0\rho_0}{2}\frac{\rho_1}{\tau_1}(\mathrm{Id}-\Omega_0\otimes\Omega_0)\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0(\omega\cdot\Omega_0)d\omega\Omega_0,\end{aligned}$$

Similarly, we have for the projection of S_1 (just change 1 to 0 and 0 to 1):

$$\begin{aligned}(\mathrm{Id}-\Omega_1\otimes\Omega_1)S_1 &= -\alpha\frac{c_1\rho_1}{2}\frac{\rho_0}{\tau_0}\langle(\sin\theta)^2h_1\rangle_{M_{\Omega_1}}\rho_0\rho_1(\mathrm{Id}-\Omega_1\otimes\Omega_1)\Omega_0\\ &\quad +(1+\alpha\frac{\rho_1}{2})\frac{\rho_0}{\tau_0}\int_{\omega\in\mathbb{S}^1} M_{\Omega_0}h_1(\omega\cdot\Omega_1)(\omega\cdot\Omega_1^\perp)d\omega\Omega_1^\perp \qquad \text{(II.69)}\\ &\quad +\frac{\alpha c_1\rho_1}{2}\frac{\rho_0}{\tau_0}(\mathrm{Id}-\Omega_1\otimes\Omega_1)\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_1(\omega\cdot\Omega_1)d\omega\Omega_1.\end{aligned}$$

We now detail the proof of proposition 4.

Proof. Let us introduce ϕ the angle between Ω_0 and Ω_1. We have the following computations.

1. We can integrate expressions between 0 and π instead of $-\pi$ and π (it will be useful

since h_0 and h_1 have constant sign on the interval $(0,\pi)$):

$$\begin{aligned}\int_{\omega\in\mathbb{S}^1} M_{\Omega_1} h_0(\omega\cdot\Omega_0)(\omega\cdot\Omega_0^\perp)d\omega &= \\ &= C_1\int_{\theta=0}^{\pi}\sin\theta(e^{\frac{1}{\lambda_1}\cos(\theta-\phi)}-e^{\frac{1}{\lambda_1}\cos(\theta+\phi)})h_0(\cos\theta)d\theta \\ &= C_1\int_{\theta=0}^{\pi}\sin\theta e^{\frac{1}{\lambda_1}\cos\theta\cos\phi}(e^{\frac{1}{\lambda_1}\sin\theta\sin\phi}-e^{-\frac{1}{\lambda_1}\sin\theta\sin\phi})h_0(\cos\theta)d\theta \\ &= 2C_1\int_{\theta=0}^{\pi}\sin\theta e^{\frac{1}{\lambda_1}\cos\theta\cos\phi}\sinh\left(\frac{1}{\lambda_1}\sin\theta\sin\phi\right)h_0(\cos\theta)d\theta,\end{aligned}$$

where C_1 is the normalization constant of the $\theta\to e^{\cos\theta/\lambda_1}$.

2. We simplify the last terms of the two equilibrium equations (II.68)-(II.69):

$$\begin{aligned}(\mathrm{Id}-\Omega_0\otimes\Omega_0)\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0(\omega\cdot\Omega_0)d\omega\Omega_0 &= \\ = (\mathrm{Id}-\Omega_0\otimes\Omega_0)\Big[\int_{\theta=-\pi}^{\pi}(\cos\theta)^2C_1e^{\frac{1}{\lambda_1}\cos(\theta-\phi)}h_0(\cos\theta)d\theta\ (\Omega_0\otimes\Omega_0)\Omega_0 \\ +\int_{\theta=-\pi}^{\pi}\sin\theta\cos\theta C_1e^{\frac{1}{\lambda_1}\cos(\theta-\phi)}h_0(\cos\theta)d\theta\left[(\Omega_0^\perp\otimes\Omega_0)+(\Omega_0\otimes\Omega_0^\perp)\right]\Omega_0 \\ +\int_{\theta=-\pi}^{\pi}(\sin\theta)^2C_1e^{\frac{1}{\lambda_1}\cos(\theta-\phi)}h_0(\cos\theta)d\theta\ (\Omega_0^\perp\otimes\Omega_0^\perp)\Omega_0\Big] \\ =\int_{\theta=-\pi}^{\pi}\sin\theta\cos\theta C_1e^{\frac{1}{\lambda_1}\cos(\theta-\phi)}h_0(\cos\theta)d\theta\ \Omega_0^\perp \\ =2C_1\int_{\theta=0}^{\pi}\sin\theta\cos\theta e^{\frac{1}{\lambda_1}\cos\theta\cos\phi}\sinh\left(\frac{1}{\lambda_1}\sin\theta\sin\phi\right)h_0(\cos\theta)d\theta\ \Omega_0^\perp.\end{aligned}$$

So, we obtain from (II.68):

$$\begin{aligned}&2C_1\int_{\theta=0}^{\pi}\left(\frac{\rho_1}{\tau_1}+\frac{\alpha\rho_0\rho_1}{2\tau_1}(1+c_0\cos\theta)\right)\sin\theta e^{\frac{1}{\lambda_1}\cos\theta\cos\phi}\sinh\left(\frac{1}{\lambda_1}\sin\theta\sin\phi\right)h_0(\cos\theta)d\theta \\ &-\frac{\alpha c_1}{2\tau_0}\beta_0\rho_0\rho_1\sin\phi=0.\end{aligned}$$

Since $\psi_0(\theta)=h_0(\cos\theta)\sin\theta$ and $\psi_1(\theta)=h_1(\cos\theta)\sin\theta$ are non positive on $(0,\pi)$ (application of the maximum principle for the elliptic equations (II.28)), the two terms of the last equations have the same sign (the sign of $\sin\phi$) and thus the two equal zero. Therefore, the angle ϕ equals 0 or π, which implies $\Omega_0=\Omega_1$ or $\Omega_0=-\Omega_1$. Similarly, equation (II.69) leads to:

$$\begin{aligned}&2C_0\int_{\theta=0}^{\pi}\left(\frac{\rho_0}{\tau_0}+\frac{\alpha\rho_1\rho_0}{2\tau_0}(1+c_1\cos\theta)\right)\sin\theta e^{\frac{1}{\lambda_0}\cos\theta\cos\phi}\sinh\left(\frac{1}{\lambda_0}\sin\theta\sin\phi\right)h_1(\cos\theta)d\theta \\ &-\frac{\alpha c_0}{2\tau_1}\beta_1\rho_0\rho_1\sin\phi=0,\end{aligned}$$

which gives the same conclusions as before. ∎

C Appendix: Proof of proposition 6 (The linearized exchange operator)

Proof. Let us begin by developing the expressions of X_0 and Y_0 introduced in lemma 5:

$$\begin{aligned}
X_0 &= \alpha\frac{\tilde{\rho}_0}{2}\rho_1\left[\int_{\omega\in\mathbb{S}^1}\omega M_{\Omega_1}h_0 d\omega\right] + \left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0 d\omega\right]\ (\alpha\frac{c_0\tilde{\rho}_0}{2}\rho_1\Omega_0 + \alpha\frac{c_0\rho_0}{2}\rho_1\tilde{\Omega}_0)\\
&+\left(1+\alpha\frac{\rho_0}{2}\right)\tilde{\rho}_1\left[\int_{\omega\in\mathbb{S}^1}\omega M_{\Omega_1}h_0 d\omega\right]\\
&+\left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0 d\omega\right]\ \left(\alpha\frac{c_0\rho_0}{2}\tilde{\rho}_1\Omega_0 + (1+\alpha\frac{\rho_0}{2})\lambda_1^{-1}\rho_1\tilde{\Omega}_1\right)\\
&+\left[\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_1)(\omega\otimes\omega)M_{\Omega_1}h_0 d\omega\right]\ \alpha\frac{c_0\rho_0}{2}\lambda_1^{-1}\rho_1\Omega_0\\
&+(1+\alpha\frac{\rho_0}{2})\rho_1\left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_1}h_0' d\omega\right]\tilde{\Omega}_0\\
&+\alpha\frac{c_0\rho_0}{2}\rho_1\left[\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_0)(\omega\otimes\omega)M_{\Omega_1}h_0' d\omega\right]\ \Omega_0,
\end{aligned}$$

$$\begin{aligned}
Y_0 &= \alpha\frac{\tilde{\rho}_1}{2}\rho_0\left[\int_{\omega\in\mathbb{S}^1}\omega M_{\Omega_0}h_0 d\omega\right] + \left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_0 d\omega\right]\ (\alpha\frac{c_1\tilde{\rho}_1}{2}\rho_0\Omega_1 + \alpha\frac{c_1\rho_1}{2}\rho_0\tilde{\Omega}_1)\\
&+\left(1+\alpha\frac{\rho_1}{2}\right)\tilde{\rho}_0\left[\int_{\omega\in\mathbb{S}^1}\omega M_{\Omega_0}h_0 d\omega\right]\\
&+\left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_0 d\omega\right]\ \left(\alpha\frac{c_1\rho_1}{2}\tilde{\rho}_0\Omega_1 + (1+\alpha\frac{\rho_1}{2})\lambda_0^{-1}\rho_0\tilde{\Omega}_0\right)\\
&+\left[\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_0)(\omega\otimes\omega)M_{\Omega_0}h_0 d\omega\right]\ \alpha\frac{c_1\rho_1}{2}\lambda_0^{-1}\rho_0\Omega_1\\
&+(1+\alpha\frac{\rho_1}{2})\rho_0\left[\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_0}h_0' d\omega\right]\tilde{\Omega}_0\\
&+\alpha\frac{c_1\rho_1}{2}\rho_0\left[\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_0)(\omega\otimes\omega)M_{\Omega_0}h_0' d\omega\right]\ \Omega_1.
\end{aligned}$$

With the assumption of the equilibrium $\Omega_0 = \Omega_1$ and the following identities:

$$\begin{aligned}
&\int_{\omega\in\mathbb{S}^1}\omega M_{\Omega_{0,1}}h_0 d\omega = \langle\cos\theta h_0\rangle_{M_{0,1}},\\
&\int_{\omega\in\mathbb{S}^1}(\omega\otimes\omega)M_{\Omega_{0,1}}h_0 d\omega = \langle\cos^2\theta h_0\rangle_{M_{0,1}}\Omega_0\otimes\Omega_0 + \langle\sin^2\theta h_0\rangle_{M_{0,1}}\Omega_0^\perp\otimes\Omega_0^\perp,\\
&\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_1)(\omega\otimes\omega)M_{\Omega_1}h_0 d\omega = (\tilde{\Omega}_1\cdot\Omega_0^\perp)\langle\sin^2\theta\cos\theta h_0\rangle_{M_1}\left(\Omega_0\otimes\Omega_0^\perp + \Omega_0^\perp\otimes\Omega_0\right),\\
&\int_{\omega\in\mathbb{S}^1}(\omega\cdot\tilde{\Omega}_0)(\omega\otimes\omega)M_{\Omega_{0,1}}h_0 d\omega = (\tilde{\Omega}_0\cdot\Omega_0^\perp)\langle\sin^2\theta\cos\theta h_0\rangle_{M_{0,1}}\left(\Omega_0\otimes\Omega_0^\perp + \Omega_0^\perp\otimes\Omega_0\right),
\end{aligned}$$

S_0 can be expressed as follows:

$$\begin{aligned}
S_0 &= \frac{1}{\tau_1}\left((1+\alpha\frac{\rho_0}{2})\rho_1\langle\cos\theta h_0\rangle_{M_1} + \alpha\frac{c_0\rho_0}{2}\rho_1\langle\cos^2\theta h_0\rangle_{M_1}\right)\Omega_0\\
&-\frac{1}{\tau_0}\left((1+\alpha\frac{\rho_1}{2})\rho_0\langle\cos\theta h_0\rangle_{M_0} + \alpha\frac{c_1\rho_1}{2}\rho_0\langle\cos^2\theta h_0\rangle_{M_0}\right)\Omega_0,
\end{aligned}$$

and X_0 and Y_0 are given by:

$$\begin{aligned}
X_0 &= \tilde{\rho}_0 \left(\frac{\alpha}{2}\rho_1\langle\cos\theta h_0\rangle_{M_1} + \frac{\alpha c_0}{2}\rho_1\langle\cos^2\theta h_0\rangle_{M_1}\right)\Omega_0 \\
&\quad +\tilde{\rho}_1 \left((1+\alpha\frac{\rho_0}{2})\langle\cos\theta h_0\rangle_{M_1} + \alpha\frac{c_0\rho_0}{2}\langle\cos^2\theta h_0\rangle_{M_1}\right)\Omega_0 \\
&\quad + \Big(\alpha\frac{c_0\rho_0}{2}\rho_1\langle\sin^2\theta h_0\rangle_{M_1} + (1+\alpha\frac{\rho_0}{2})\rho_1\langle\sin^2\theta h_0'\rangle_{M_1} \\
&\qquad +\alpha\frac{\rho_0 c_0}{2}\rho_1\langle\sin^2\theta\cos\theta h_0'\rangle_{M_1}\Big)\tilde{\Omega}_0 \\
&\quad + \left((1+\alpha\frac{\rho_0}{2})\lambda_1^{-1}\rho_1\langle\sin^2\theta h_0\rangle_{M_1} + \alpha\frac{c_0\rho_0}{2}\rho_1\lambda_1^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_1}\right)\tilde{\Omega}_1,
\end{aligned}$$

$$\begin{aligned}
Y_0 &= \tilde{\rho}_0 \left((1+\alpha\frac{\rho_1}{2})\langle\cos\theta h_0\rangle_{M_0} + \alpha\frac{c_1\rho_1}{2}\langle\cos^2\theta h_0\rangle_{M_0}\right)\Omega_0 \\
&\quad +\tilde{\rho}_1 \left(\frac{\alpha}{2}\rho_0\langle\cos\theta h_0\rangle_{M_0} + \frac{\alpha c_1}{2}\rho_0\langle\cos^2\theta h_0\rangle_{M_0}\right)\Omega_0 \\
&\quad + \Big((1+\alpha\frac{\rho_1}{2})\lambda_0^{-1}\rho_0\langle\sin^2\theta h_0\rangle_{M_0} + \alpha\frac{c_1\rho_1}{2}\rho_0\lambda_0^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_0} \\
&\qquad +(1+\alpha\frac{\rho_1}{2})\rho_0\langle\sin^2\theta h_0'\rangle_{M_0} + \alpha\frac{c_1\rho_1}{2}\rho_0\langle\sin^2\theta\cos\theta h_0'\rangle_{M_0}\Big)\tilde{\Omega}_0 \\
&\quad + \left(\alpha\frac{c_1\rho_1}{2}\rho_0\langle\sin^2\theta h_0\rangle_{M_0}\right)\tilde{\Omega}_1.
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
&(\mathrm{Id}-\Omega_0\otimes\Omega_0)\tilde{S}_0 - (\tilde{\Omega}_0\cdot S_0)\Omega_0 - (\Omega_0\cdot S_0)\tilde{\Omega}_0 \\
&\quad = \frac{1}{\tau_1}(1+\alpha\frac{\rho_0}{2})\rho_1\left[\lambda_1^{-1}\langle\sin^2\theta h_0\rangle_{M_1}\tilde{\Omega}_1 + \langle\sin^2\theta h_0'\rangle_{M_1}\tilde{\Omega}_0 - \langle\cos\theta h_0\rangle_{M_1}\tilde{\Omega}_0\right] \\
&\qquad +\frac{1}{\tau_1}\alpha\frac{c_0\rho_0}{2}\rho_1\Big[\langle\sin^2\theta h_0\rangle_{M_1}\tilde{\Omega}_0 + \langle\sin^2\theta\cos\theta h_0'\rangle_{M_1}\tilde{\Omega}_0 \\
&\qquad\qquad +\lambda_1^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_1}\tilde{\Omega}_1 - \langle\cos^2\theta h_0\rangle_{M_1}\tilde{\Omega}_0\Big] \\
&\qquad -\frac{1}{\tau_0}(1+\alpha\frac{\rho_1}{2})\rho_0\left[\lambda_0^{-1}\langle\sin^2\theta h_0\rangle_{M_0}\tilde{\Omega}_0 + \langle\sin^2\theta h_0'\rangle_{M_0}\tilde{\Omega}_0 - \langle\cos\theta h_0\rangle_{M_0}\tilde{\Omega}_0\right] \\
&\qquad -\frac{1}{\tau_0}\alpha\frac{c_1\rho_1}{2}\rho_0\Big[\langle\sin^2\theta h_0\rangle_{M_0}\tilde{\Omega}_1 + \langle\sin^2\theta\cos\theta h_0'\rangle_{M_0}\tilde{\Omega}_0 \\
&\qquad\qquad +\lambda_0^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_0}\tilde{\Omega}_0 - \langle\cos^2\theta h_0\rangle_{M_0}\tilde{\Omega}_0\Big], \\
&\quad = A_0(\tilde{\Omega}_1 - \tilde{\Omega}_0),
\end{aligned}$$

with

$$\begin{aligned}
A_0 &= \frac{1}{\tau_1}(1+\alpha\frac{\rho_0}{2})\rho_1\lambda_1^{-1}\langle\sin^2\theta h_0\rangle_{M_1} \\
&\quad +\frac{1}{\tau_1}\alpha\frac{c_0\rho_0}{2}\rho_1\lambda_1^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_1} - \frac{1}{\tau_0}\alpha\frac{c_1\rho_1}{2}\rho_0\langle\sin^2\theta h_0\rangle_{M_0},
\end{aligned}$$

where we used the identities:

$$\begin{aligned}
&\lambda_0^{-1}\langle\sin^2\theta h_0\rangle_{M_0} = \langle\cos\theta h_0\rangle_{M_0} - \langle\sin^2\theta h_0'\rangle_{M_0}, \\
&\lambda_0^{-1}\langle\sin^2\theta\cos\theta h_0\rangle_{M_0} = \langle\cos^2\theta h_0\rangle_{M_0} - \langle\sin^2\theta h_0\rangle_{M_0} - \langle\sin^2\theta\cos\theta h_0'\rangle_{M_0}.
\end{aligned}$$

We also used the fact that S_0 is parallel to Ω_0 and thus the term $(\tilde{\Omega}_0 \cdot S_0)\Omega_0$ vanishes. We similarly obtain:

$$(\mathrm{Id} - \Omega_1 \otimes \Omega_1)\tilde{S}_1 - (\tilde{\Omega}_1 \cdot S_1)\Omega_1 - (\Omega_1 \cdot S_1)\tilde{\Omega}_1 = A_1(\tilde{\Omega}_0 - \tilde{\Omega}_1),$$

with

$$\begin{aligned} A_1 &= \frac{1}{\tau_0}(1 + \alpha\frac{\rho_1}{2})\rho_0\lambda_0^{-1}\langle \sin^2\theta h_1\rangle_{M_1} + \frac{1}{\tau_0}\alpha\frac{c_1\rho_1}{2}\rho_0\lambda_0^{-1}\langle \sin^2\theta\cos\theta h_1\rangle_{M_0} \\ &\quad -\frac{1}{\tau_1}\alpha\frac{c_0\rho_0}{2}\rho_1\langle \sin^2\theta h_1\rangle_{M_1}. \end{aligned}$$

■

Bibliography

[1] P. Degond. Macroscopic limits of the Boltzmann equation: a review. In P. Degond, L. Pareschi, and R. Russo, editors, *Modeling and computational methods for kinetic equations*, volume Modeling and Simulation in Science of *Engineering and Technology Series*, pages 3–57. Birkhaüser, Boston, 2003.

[2] P. Degond and S. Motsch. Continuum limit of self-driven particles with orientation interaction. *Math. Models Methods Appl. Sci.*, 18:1193–1215, 2008.

[3] M.H. Pillot, J. Gautrais, J. Gouello, P. Michelena, A. Sibbald, and R. Bon. Moving together: incidental leaders and naïve followers. *Behav. Proc.*, 83(3):235–241, 2010.

[4] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel Type of Phase Transition in a System of Self-Driven Particles. *Phys. Rev. Lett.*, 75(6):1226–1229, 1995.

Chapitre III

Simulations numériques du système de Vicsek macroscopique

Ce chapitre a été écrit en collaboration avec Sébastien Motsch. Il fait l'objet d'un article intitulé : ***Numerical simulations of a non-conservative hyperbolic system with geometric constraints describing swarming behaviour****. La suite de ce chapitre est écrite en anglais.*

Numerical simulations of a non-conservative hyperbolic system with geometric constraints describing swarming behaviour

Abstract: The Vicsek model is a very popular individual based model which describes collective behavior among animal societies. A macroscopic version of the Vicsek model has been derived from a large scale limit of this individual based model [12]. In this work, we want to numerically validate this Macroscopic Vicsek model (MV). To this aim, we compare the simulations of the macroscopic and microscopic models one with each other. The MV model is a non-conservative hyperbolic equation with a geometric constraint. Due to the lack of theory for this kind of equations, we derive several equivalents for this system leading to specific numerical schemes. The numerical simulations reveal that the microscopic and macroscopic models are in good agreement provided that we choose one of the proposed formulations based on a relaxation of the geometric constraint. This confirms the relevance of the macroscopic equation but it also calls for a better theoretical understanding of this type of equations.

1 Introduction

This paper is devoted to the numerical study of a macroscopic version of the Vicsek model which describes swarming behavior. This macroscopic model has been derived in [12] from the microscopic Vicsek model [26]. The goal of this work is to provide a numerical validation of the macroscopic model by comparing it with simulations of the microscopic model.

The Vicsek model [26] is widely used to describe swarming behavior such as flock of birds [2], schools of fish [1, 9, 19, 25] (in this case the model is combined with an

attractive-repulsive force) or recently the motion of locusts [5]. In this model, individuals have a constant velocity and they tend to align with their neighbors. Despite the simplicity of the model, a lot of questions remain open about it. A first field of research concerns phase transitions within the model depending on the level of noise [6, 16, 24, 26]. Another question arises from the long time dynamics of the model [10, 11, 17]: is there convergence to a stationary state of the system? From another perspective, since collective displacements in natural environment can concern up to several million individuals, it is natural to look for a macroscopic version of the Vicsek model. On the one hand, macroscopic models constitute powerful analytical tools to study the dynamics at large scales [8, 13, 23]. On the other hand, the related numerical schemes are computationally much more efficient compared with particle simulations of a large number of interacting agents. In [12], a Macroscopic Vicsek model (MV) has been derived from a large scale limit of the microscopic Vicsek model. The macroscopic model is obtained from a rigorous perturbation theory of the original Vicsek model. Another macroscopic model is obtained in [3] based on more phenomenological closure assumptions.

The MV model presents several specificities which make the model interesting. First, it is a non-conservative hyperbolic system and secondly it involves a geometric constraint. These are the consequences at the macroscopic level of two specificities of the microscopic model: the total momentum is not conserved by the particle dynamics and the speed of the particles is constant. The first property is an intrinsic property of self-propelled particles and the second property is a usual assumption in the models of collective displacements [9, 15, 26]. Up to our knowledge the theory of such systems is almost empty. Non-conservative systems have been studied in the literature [4, 7, 20] but none of them involve geometric constraints.

In this work, since a theoretical framework for such systems is not available, we adopt several approaches. First, we introduce a conservative formulation for the 1D formulation of the MV model which is equivalent to the initial one for smooth solutions only. With this conservative formulation, we can use standard hyperbolic theory to build Riemann problem solution and shock capturing schemes [21]. The numerical scheme based on the conservative formulation is called *conservative method*. But since the equivalence with the original formulation is only valid for smooth solutions [22], there is no guarantee that the conservative formulation gives the right answer at shocks. For this reason, we introduce another formulation of the MV model where the constraint is treated through the relaxation limit of an unconstrained conservative system. This formulation leads to a natural numerical scheme based on a splitting between the conservative part of the equation and the relaxation. This scheme will be referred to as the *splitting method*. For comparison purposes, two other numerical schemes are also used, an *upwind scheme* and a *semi-conservative* one (where only the mass conservation equation is treated in a conservative way).

The numerical simulations of the MV model reveal that the numerical schemes all agree on rarefaction waves but disagree on shock waves. To determine the correct solution, we use the microscopic model in a regime where its solution is close to that of the macroscopic model. In practice, this corresponds to regimes where the number of particles per domain of interaction is high. The splitting method turns out to be in good agreement with particle simulations of the microscopic model, by contrast with the other schemes. In particular, for an initial condition with a contact discontinuity, the solution given by

the conservative form is simply a convection of the initial condition whereas the splitting method and the particle simulations agree on a different and more complex solution.

These results show first that the MV model well describes the microscopic model in the dense regime. Secondly, that the correct formulation of the MV model is given by the limit of a conservative equation with a stiff relaxation term.

The theoretical and numerical studies of the MV model highlight the specificity of non-conservative hyperbolic models with geometric constraints. More theoretical work is necessary in order to understand why the splitting method matches the microscopic model whereas the other methods do not. In particular, an extension of the theory developed in [7] to non-conservative relaxed models would be highly desirable.

The outline of the paper is as follows: first, we present the Vicsek and MV models in section 2. Then, we analyze the MV model and give two different formulations of the model in section 3. We develop different numerical schemes based on these formulations and we use them to numerically solve different Riemann problems in section 4. Finally, we compare simulations of the microscopic model with those of the macroscopic system in the same situations in section 5. Finally, we draw a conclusion.

2 Presentation of the Vicsek and Macroscopic Vicsek models

At the particle level, the Vicsek model describes the motion of particles which tend to align with their neighbors. We denote by x_k the position vector of the k^{th} particle and by ω_k its velocity with a constant speed ($|\omega_k| = 1$). To simplify, we suppose that the particles move in a plane. Therefore $x_k \in \mathbb{R}^2$ and $\omega_k \in S^1$. The Vicsek model at the microscopic level is given by the following equations (in dimensionless variables):

$$\frac{dx_k}{dt} = \omega_k, \tag{III.1}$$

$$d\omega_k = (\mathrm{Id} - \omega_k \otimes \omega_k)(\bar{\omega}_k\, dt + \sqrt{2d}\, dB_t), \tag{III.2}$$

where Id is the identity matrix and the symbol $\otimes$ denotes the tensor product of vectors. d is the intensity of noise, B_t is the Brownian motion and $\bar{\omega}_k$ is the direction of mean velocity around the k$^{\text{th}}$ particle defined by:

$$\bar{\omega}_k = \frac{J_k}{|J_k|}\,, \qquad J_k = \sum_{j,\,|x_j - x_k| \le R} \omega_j, \tag{III.3}$$

where R defines the radius of the interaction region. Equation (III.2) expresses the tendency of particles to move in the same direction as their neighbors. The operator $(\mathrm{Id} - \omega_k \otimes \omega_k)$ is the orthogonal projector onto the plane perpendicular to ω_k. It ensures that the speed of particles remains constant. This model is already a modification of the original Vicsek model [26], which is a time-discrete algorithm.

The Macroscopic Vicsek model (MV) describes the evolution of two macroscopic quantities: the density of particles ρ and the direction of the flow Ω. The evolution of ρ and

Ω is governed by the following equations:

$$\partial_t \rho + \nabla_x \cdot (c_1 \rho \Omega) = 0, \tag{III.4}$$
$$\rho \left(\partial_t \Omega + c_2 (\Omega \cdot \nabla_x)\Omega\right) + \lambda \left(\mathrm{Id} - \Omega \otimes \Omega\right) \nabla_x \rho = 0, \tag{III.5}$$
$$|\Omega| = 1, \tag{III.6}$$

where c_1, c_2 and λ are some constants depending on the noise parameter d. The expressions of c_1, c_2 and λ are given in appendix A. By contrast with the standard Euler system, the two convection coefficients c_1 and c_2 are different. The other specificity of this model is the constraint $|\Omega| = 1$. The operator $(\mathrm{Id} - \Omega \otimes \Omega)$ ensures that this constraint is propagated provided that it is true at the initial time. The passage from (III.1)-(III.2) to (III.4)-(III.5)-(III.6) is detailed in [12]. We note that vortex configurations are special stationary solutions of this model in two dimensions (see appendix B). Up to our knowledge, this is the first swarming model which have such analytical solutions.

3 The Macroscopic Vicsek model

3.1 Theoretical analysis of the macroscopic model

To study model (III.4)-(III.5)-(III.6), we first use the rescaling $x' = x/c_1$. Then equations (III.4)-(III.5)-(III.6) are written:

$$\partial_t \rho + \nabla_{x'} \cdot (\rho \Omega) = 0, \tag{III.7}$$
$$\rho \left(\partial_t \Omega + c' (\Omega \cdot \nabla_{x'})\Omega\right) + \lambda' \left(\mathrm{Id} - \Omega \otimes \Omega\right) \nabla_{x'} \rho = 0, \tag{III.8}$$
$$|\Omega| = 1, \tag{III.9}$$

with $c' = c_2/c_1$ and $\lambda' = \lambda/c_1$. In the sequel, we drop the primes for clarity. We refer to the appendix A for the computation of c and λ and we just mention that we have (see figure III.17):

$$\frac{1}{2} < c < 1 \quad \text{and} \quad \lambda > 0, \quad \text{for all } d > 0. \tag{III.10}$$

In two dimensions, we can use a parameterization of Ω in polar coordinates: $\Omega = (\cos\theta, \sin\theta)^T$. Therefore, equations (III.7)-(III.8) can be rewritten as:

$$\partial_t \rho + \partial_x \left(\rho \cos\theta\right) + \partial_y \left(\rho \sin\theta\right) = 0, \tag{III.11}$$
$$\partial_t \theta + c \cos\theta \partial_x \theta + c \sin\theta \partial_y \theta + \lambda \left(-\frac{\sin\theta}{\rho} \partial_x \rho + \frac{\cos\theta}{\rho} \partial_y \rho \right) = 0. \tag{III.12}$$

In this section, we suppose that ρ and θ are independent of y meaning that we are looking at waves which propagate in the x-direction. Under this assumption, the system reads:

$$\partial_t \begin{pmatrix} \rho \\ \theta \end{pmatrix} + A(\rho, \theta)\, \partial_x \begin{pmatrix} \rho \\ \theta \end{pmatrix} = 0, \tag{III.13}$$

with

$$A(\rho, \theta) = \begin{bmatrix} \cos\theta & -\rho \sin\theta \\ -\frac{\lambda \sin\theta}{\rho} & c \cos\theta \end{bmatrix}. \tag{III.14}$$

The characteristic velocities of this system are given by

$$\gamma_{1,2} = \frac{1}{2}\left[(c+1)\cos\theta \pm \sqrt{(c-1)^2\cos^2\theta + 4\lambda\sin^2\theta}\right] \tag{III.15}$$

with $\gamma_1 < \gamma_2$. Therefore, the system is strictly *hyperbolic*. A possible choice of right eigenvectors is

$$\vec{r}_1 = \begin{pmatrix} \rho\sin\theta \\ \cos\theta - \gamma_1 \end{pmatrix} \quad , \quad \vec{r}_2 = \begin{pmatrix} c\cos\theta - \gamma_2 \\ \frac{\lambda\sin\theta}{\rho} \end{pmatrix}. \tag{III.16}$$

The two fields are genuinely nonlinear except at $\theta = 0$, $\theta = \pi$ and at the extrema values of γ_p which satisfy:

$$\tan^2\theta = \frac{1}{4\lambda}\left[\frac{((c-1)^2 - 4\lambda)^2}{(c+1)^2} - (c-1)^2\right].$$

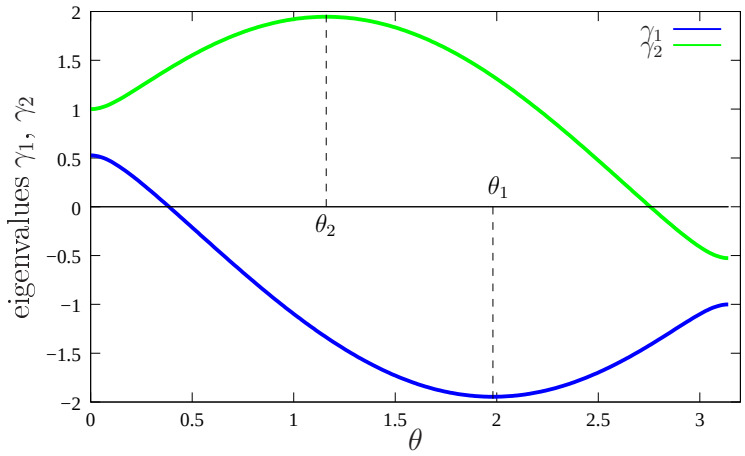


Figure III.1: The two eigenvalues γ_1 and γ_2 depending on θ ($d = 1$ in this graph). For each curve, there exists a unique extremum (θ_1 and θ_2) which corresponds to a degeneracy of the system.

The Riemann invariants of the system (III.13) are given by:

$$I_1(\rho,\theta) = \log\rho - \int_{\theta_0}^{\theta} \frac{\sin s}{\cos s - \gamma_1(s)}\, ds, \tag{III.17}$$

$$I_2(\rho,\theta) = \log\rho - \int_{\theta_0}^{\theta} \frac{c\cos s - \gamma_2(s)}{\lambda\sin s}\, ds. \tag{III.18}$$

The integral curve w_1 and w_2 starting from (ρ_l, θ_l) are given by:

$$\rho_1(\theta) = \rho_l \exp\left(\int_{\theta_0}^{\xi} \frac{\sin s}{\cos s - \gamma_1(s)}\, ds\right), \tag{III.19}$$

$$\rho_2(\theta) = \rho_l \exp\left(\int_{\theta_0}^{\xi} \frac{c\cos s - \gamma_2(s)}{\lambda\sin s}\, ds\right). \tag{III.20}$$

These are the rarefaction curves. To select the physically admissible rarefaction curve, we remark that γ_p must grow from the left to right states. The proofs of these elementary facts are omitted. The quantities $\gamma_{1,2}$ as functions of θ are depicted in figure III.1.

3.2 A conservative form of the MV model in dimension 1

For non-conservative systems, shock waves are not uniquely defined [20, 22]. However, in the present case, a conservative formulation of the system can be found in dimension 1. Indeed, it is an easy matter to see that, if $\sin\theta \neq 0$, system (III.13) can be rewritten in conservative form:

$$\partial_t \begin{pmatrix} \rho \\ f_1(\theta) \end{pmatrix} + \partial_x \begin{pmatrix} \rho\cos\theta \\ cf_2(\theta) - \lambda\log(\rho) \end{pmatrix} = 0, \tag{III.21}$$

with:

$$f_1(\theta) = \log\left|\tan\frac{\theta}{2}\right| = \log\left|\frac{\sin\theta}{\cos\theta+1}\right|, \tag{III.22}$$

$$f_2(\theta) = \log|\sin\theta|\,. \tag{III.23}$$

However, the functions f_1 and f_2 are singular when $\sin\theta = 0$ which means that the conservative form is only valid as long as θ stays away from $\theta = 0$.

The conservative form (III.21) leads to the following Rankine-Hugoniot conditions for shock waves: two states (ρ_l, θ_l) and (ρ_r, θ_r) are connected by a shock wave traveling at a constant speed s if they satisfy:

$$s\begin{pmatrix} \rho_r - \rho_l \\ f_1(\theta_r) - f_1(\theta_l) \end{pmatrix} = \begin{pmatrix} \rho_r\cos\theta_r - \rho_l\cos\theta_l \\ cf_2(\theta_r) - cf_2(\theta_l) - \lambda\log\rho_r + \lambda\log\rho_l \end{pmatrix}. \tag{III.24}$$

We can combine the two equations of the system (III.24) to eliminate the constant s and this leads to the following expression of the shock curve:

$$\begin{aligned}(\rho_r - \rho_l)(cf_2(\theta_r) - cf_2(\theta_l) - \lambda\log\rho_r + \lambda\log\rho_l) \\ = (\rho_r\cos\theta_r - \rho_l\cos\theta_l)(f_1(\theta_r) - f_1(\theta_l)).\end{aligned} \tag{III.25}$$

This equation must be numerically solved. The entropic part of the shock curve is determined by the requirement that γ_p must satisfy the Lax entropy condition. In figure III.2, we give an example of a solution of a Riemann problem obtained by computing the intersection of the shock and rarefaction curves.

3.3 The MV model as the relaxation limit of a conservative system

We are going to prove that the MV model (III.7)-(III.8)-(III.9) can be seen as the relaxation limit of a conservative hyperbolic model with a relaxation term. This link will be used later to build a new numerical scheme. More precisely, we introduce the relaxation model:

$$\partial_t\rho^\varepsilon + \nabla_x\cdot(\rho^\varepsilon\Omega^\varepsilon) = 0, \tag{III.26}$$

$$\partial_t(\rho^\varepsilon\Omega^\varepsilon) + c\nabla_x\cdot(\rho^\varepsilon\Omega^\varepsilon\otimes\Omega^\varepsilon) + \lambda\nabla_x\rho^\varepsilon = \frac{\rho^\varepsilon}{\varepsilon}(1-|\Omega^\varepsilon|^2)\Omega^\varepsilon. \tag{III.27}$$

In this model, the constraint $|\Omega| = 1$ is replaced by a relaxation operator. Formally, in the limit $\varepsilon \to 0$, we recover the constraint $|\Omega| = 1$.

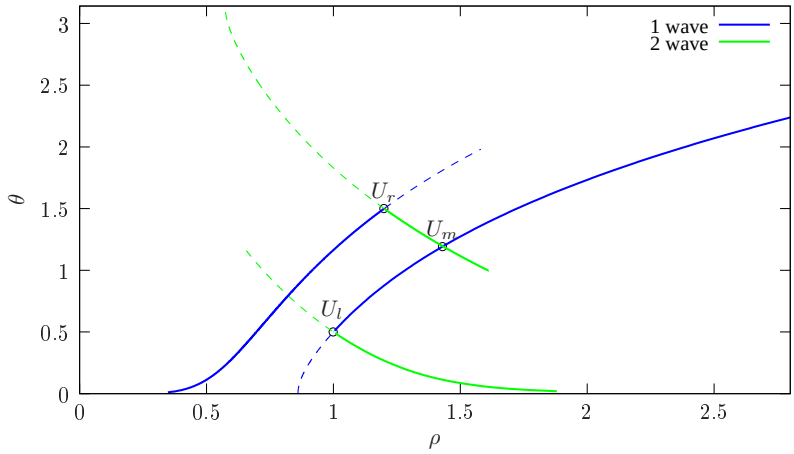


Figure III.2: A solution of the Riemann problem with left and right states U_l and U_r (solid line for shock waves and dotted line for rarefaction waves). In this example, the solution is given by two shock waves.

Proposition 8. *The relaxation model (III.26)-(III.26) converges to the MV model (III.7)-(III.8)-(III.9) as ε goes to zero.*

Proof (formal). We define $R^\varepsilon = \rho^\varepsilon(1 - |\Omega^\varepsilon|^2)\Omega^\varepsilon$. Suppose that as ε goes to zero:

$$\rho^\varepsilon \overset{\varepsilon\to 0}{\longrightarrow} \rho^0 \quad , \quad \Omega^\varepsilon \overset{\varepsilon\to 0}{\longrightarrow} \Omega^0. \tag{III.28}$$

Then $R^\varepsilon \overset{\varepsilon\to 0}{\longrightarrow} 0$, which generically implies that $|\Omega^0|^2 = 1$ (except where $\rho^0\Omega^0 = 1$ which one assumes to be a negligible set). Therefore, we have:

$$\partial_t\Omega^0 \cdot \Omega^0 = 0 \quad , \quad (\Omega^0 \cdot \nabla_x)\Omega^0 \cdot \Omega^0 = 0. \tag{III.29}$$

Then since $R^\varepsilon \times \Omega^\varepsilon = 0$, we have:

$$\left(\partial_t\left(\rho^\varepsilon\Omega^\varepsilon\right) + c\nabla_x \cdot \left(\rho^\varepsilon\Omega^\varepsilon \otimes \Omega^\varepsilon\right) + \lambda\,\nabla_x\rho^\varepsilon\right) \times \Omega^\varepsilon = 0.$$

and consequently when $\varepsilon \to 0$:

$$\partial_t\left(\rho^0\Omega^0\right) + c\nabla_x \cdot \left(\rho^0\Omega^0 \otimes \Omega^0\right) + \lambda\,\nabla_x\rho^0 = \alpha\Omega^0, \tag{III.30}$$

for a real number α to be determined. Taking the scalar product of (III.30) with Ω^0 and using (III.29), we find:

$$\alpha = \partial_t\rho^0 + c\nabla_x \cdot (\rho^0\Omega^0) + \lambda\nabla_x\rho^0 \cdot \Omega^0.$$

Using the conservation of mass ($\partial_t\rho^0 = -\nabla_x \cdot (\rho^0\Omega^0)$), we finally have:

$$\alpha = (c-1)\nabla_x \cdot (\rho^0\Omega^0) + \lambda\nabla_x\rho^0 \cdot \Omega^0.$$

Therefore, the relaxation term satisfies:

$$\frac{1}{\varepsilon}R^\varepsilon = [(c-1)\nabla_x \cdot (\rho^0\Omega^0) + \lambda\nabla_x\rho^0 \cdot \Omega^0]\Omega^0 + O(\varepsilon).$$

Inserting in (III.26)-(III.27) and taking the limit $\varepsilon \to 0$, we recover the MV model (III.7)-(III.8) at the first order in ε.

■

Remark. As for the MV model, we can also analyze the hyperbolicity of the left hand side of (III.26)-(III.27). The eigenvalues are given by:

$$\gamma_1 = cu - \sqrt{\Delta} \quad , \quad \gamma_2 = cu \quad , \quad \gamma_3 = cu + \sqrt{\Delta},$$

where u denotes the x-coordinate of Ω and $\Delta = \lambda - (c - c^2)u^2$. The system is hyperbolic if and only if $|u| < \sqrt{\frac{\lambda}{c-c^2}}$. As we can see in figure III.3, for $u^2 = 1$, Δ is positive for any values of the noise parameter d. In particular, this implies that the relaxation model is hyperbolic for every $|u| \leq 1$.

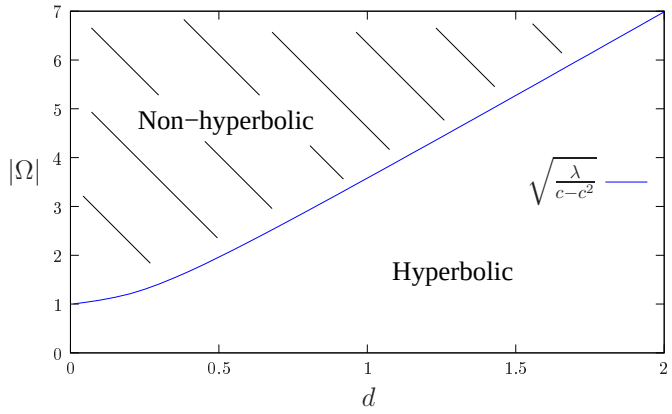


Figure III.3: The quantity $\sqrt{\lambda/(c-c^2)}$ depending on d. The relaxation model (III.26)-(III.26) is hyperbolic when the speed $|\Omega|$ is below this curve. At the limit $\varepsilon \to 0$, $|\Omega^\varepsilon| \to 1$ and therefore the relaxation model is hyperbolic for any d in this limit.

4 Numerical simulations of the MV model

4.1 Numerical schemes

We propose four different numerical schemes to solve the MV model. The first two schemes originate from the discussions of the previous section, the two other one are based on the non-conservative form of the MV model.
We use the following notations: we fix a uniform stencil $(x_i)_i$ (with $|x_{i+1} - x_i| = \Delta x$) and a time step Δt. We denote by $U_i^n = (\rho_i^n, \theta_i^n)$ the value of the mass and flux direction at the position x_i and at time $n\Delta t$.

4.1.1 The conservative scheme

Here we use the conservative form of the MV model (III.21):

$$\partial_t V + \partial_x F(V) = 0, \tag{III.31}$$

with $V = (\rho, f_1(\theta))^T$ and $F(V) = (\rho\cos\theta,\, cf_2(\theta) - \lambda\log(\rho))^T$. We use a Roe method to discretize this equation:

$$\frac{V_i^{n+1} - V_i^n}{\Delta t} + \frac{\widehat{F}_{i+} - \widehat{F}_{i-}}{\Delta x} = 0, \tag{III.32}$$

where the intermediate flux $\widehat{F}_{i+}$ is given by:

$$\widehat{F}_{i+} = \frac{F(V_i) + F(V_{i+1})}{2} - \left|\mathcal{A}(\overline{V}_{i+})\right| \frac{V_{i+1} - V_i}{2}, \tag{III.33}$$

and $\mathcal{A}$ is the Jacobian of the flux F:

$$\mathcal{A}(V) = DF(V) = \begin{bmatrix} \cos\theta & -\rho\sin^2\theta \\ -\frac{\lambda}{\rho} & c\cos\theta \end{bmatrix} \tag{III.34}$$

calculated at the mean value $\overline{V}_{i+} = \frac{V_i+V_{i+1}}{2}$.

As mentioned earlier, the conservative form is only valid when θ does not cross a singularity $\theta = 0$ or $\theta = \pi$ (i.e. $\sin\theta = 0$). Nevertheless, numerically we can still use the formulation (III.32) when θ changes sign. Moreover, since f_1 is an even function, this only gives $|\theta^{n+1}|$. To determine the sign of θ^{n+1}, we use an auxiliary value $\widehat{\theta}$ which we update with the upwind scheme (III.38). The sign of θ is then determined using the sign of $\widehat{\theta}$.

4.1.2 The splitting method

The next scheme uses the relaxation model (III.26)-(III.27). The idea is to split the relaxation model in two parts, first the conservative part:

$$\begin{aligned} &\partial_t\rho + \nabla_x\cdot(\rho\Omega) = 0, \\ &\partial_t(\rho\Omega) + c\nabla_x\cdot(\rho\Omega\otimes\Omega) + \lambda\nabla_x\rho = 0. \end{aligned} \tag{III.35}$$

and then the relaxation part:

$$\begin{aligned} \partial_t\rho &= 0, \\ \partial_t(\rho\Omega) &= \frac{\rho}{\varepsilon}(1-|\Omega|^2)\Omega. \end{aligned} \tag{III.36}$$

This system reduces to: $\partial_t\Omega = \frac{1}{\varepsilon}(1-|\Omega|^2)\Omega$. Since this equation only changes the vector field Ω in norm (i.e. $\partial_t\Omega\cdot\Omega^\perp = 0$), we can once again reduce this equation to:

$$\frac{1}{2}\partial_t|\Omega|^2 = \frac{1}{\varepsilon}(1-|\Omega|^2)|\Omega|^2. \tag{III.37}$$

Equation (III.37) can be explicitly solved: $|\Omega|^2 = (1 + C_0\,\mathrm{e}^{-2/\varepsilon\,t})^{-1}$ with $C_0 = \left(\frac{1}{|\Omega_0|^2} - 1\right)$. We indeed take the limit $\varepsilon \to 0$ of this expression and replace the relation term by a mere normalization: $\Omega \to \Omega/|\Omega|$.

The conservative part is solved by a Roe method with a Roe matrix computed following [22] page 156.

4.1.3 Non-conservative schemes

We present two other numerical schemes based on the non-conservative formulation of the MV model.

(i) Upwind scheme

The method consists to update the value of U_i^n with the formula:

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + A^+ \left(\frac{U_i^n - U_{i-1}^n}{\Delta t} \right) + A^- \left(\frac{U_{i+1}^n - U_i^n}{\Delta t} \right) = 0, \tag{III.38}$$

where A^+ and A^- are (respectively) the positive and negative part of A, defined such that $A = A^+ - A^-$ and $|A| = A^+ + A^-$ and A^+, A^- are computed using an explicit diagonalization of A.

(ii) Semi-conservative scheme

One of the problem with the upwind scheme is that it does not conserve the total mass ($\int_x \rho(x)\, dx$). In order to keep this quantity constant in time, we use the equation of conservation of mass (III.7) in a conservative form:

$$\partial_t \rho + \partial_x H(\rho, \theta) = 0, \tag{III.39}$$

with $H(\rho, \theta) = \rho \cos\theta$. Therefore, a conservative numerical scheme associated with this equation would be:

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{\widehat{H}_{i+1/2} - \widehat{H}_{i-1/2}}{\Delta x} = 0, \tag{III.40}$$

where $\widehat{H}_{i+}$ is the numerical estimation of the flux H at the interface between x_i and x_{i+1}. To estimate numerically this flux, we use the following formula with $U_i = (\rho_i, \theta_i)$:

$$\widehat{H}_{i+1/2} = H(U_{i+1/2}) - |A|_\rho \left(\frac{U_{i+1}^n - U_i^n}{2} \right), \tag{III.41}$$

where the intermediate value is given by $U_{i+1/2} = \frac{U_i^n + U_{i+1}^n}{2}$ and $|A|_\rho$ is the first line of the absolute value of A.

For the estimation of the angle θ, we keep the same scheme as for the upwind scheme. This numerical scheme uses one conservative equation (for the mass ρ) and a non-conservative equation (for the angle θ). It is thus referred to as the *semi-conservative* scheme.

4.2 Numerical simulations

To compare the various numerical schemes, we use a Riemann problem as initial condition. We choose solutions which consist of a rarefaction wave (figure III.4) or a single shock wave (figures III.5-III.6).

We take the following parameters: $d = 1$, the length of the domain is 10 units and the discontinuity for the Riemann problem is at $x = 5$ (the middle of the domain). The simulations are run during two time units with a time step $\Delta t = 2.10^{-2}$ and a space step $\Delta x = 5.10^{-2}$. For these values, the Courant number (C_n) is 0.778. We use homogeneous Neumann conditions as boundary conditions.

For the rarefaction wave, we take:

$$(\rho_l, \theta_l) = (2, 1.7) \quad , \quad (\rho_r, \theta_r) = (1.12, 0.60). \tag{III.42}$$

All the numerical schemes capture well the theoretical solution (see figure III.4).

For the shock wave, we choose:

$$(\rho_l, \theta_l) = (1, 1.05) \quad , \quad (\rho_r, \theta_r) = (1.432\,,\, 1.7). \tag{III.43}$$

For these values, the shock speed computed with (III.25) is $s = -1.585$. The results of the numerical simulations using the four schemes are given in figure III.5. The numerical solutions are in accordance with the theoretical solution given by the conservative formulation for all the numerical schemes. Nevertheless, the conservative scheme is in better accordance with this solution. For the other schemes, the shock speed differs slightly.
A second example of a shock wave is computed using the following initial condition:

$$(\rho_l, \theta_l) = (1, 0.314) \quad , \quad (\rho_r, \theta_r) = (2, 1.54). \tag{III.44}$$

The solutions given by the 4 numerical schemes are very different. Only the conservative method is in agreement with the solution given by the conservative formulation. But the conservative formulation is not necessary the right one. Indeed, in the next section, particle simulations show that the right solution is not given by the conservative formulation but rather by the splitting method.

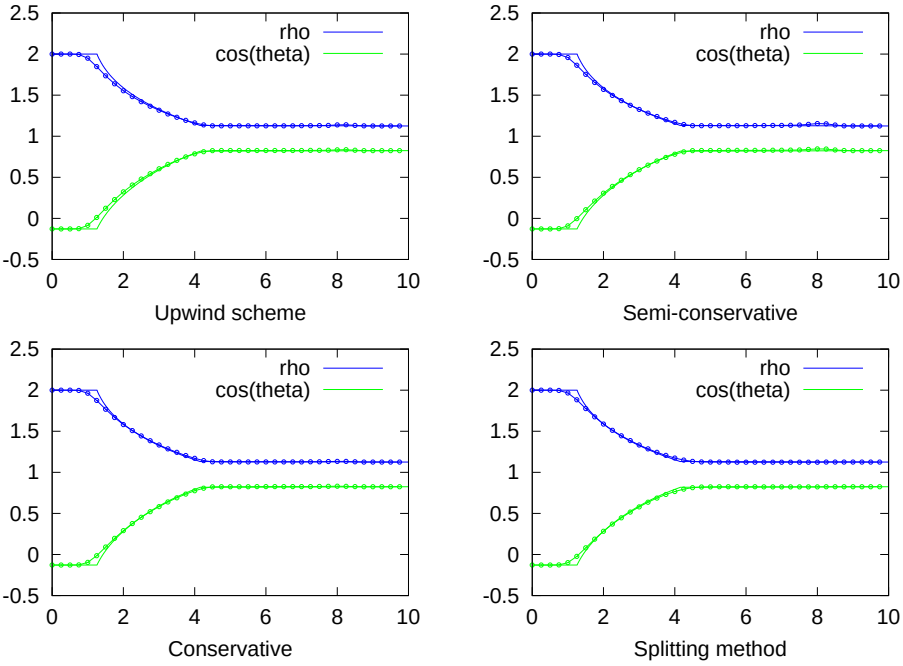


Figure III.4: The theoretical solution of the Riemann problem (III.42) given by a rarefaction curve (solid line) and the numerical solutions (points), ρ (blue) and $\cos\theta$ (green) as functions of space. The simulations are run during 2 time units, with a time step $\Delta t = 2.10^{-2}$ and a space step $\Delta x = 5.10^{-2}$ (CFL=.778).

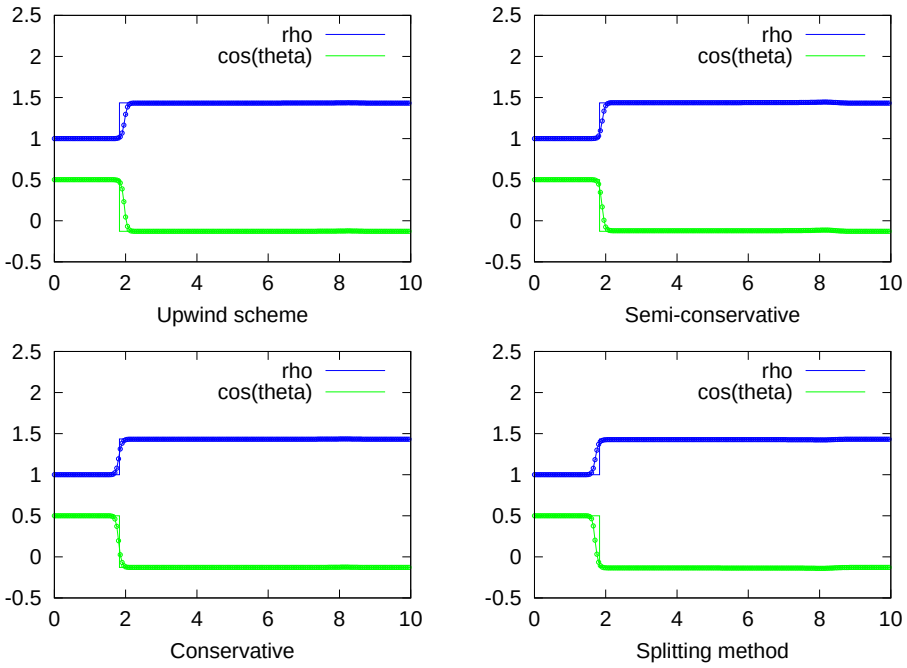


Figure III.5: Theoretical and numerical solutions of the Riemann problem (III.43)

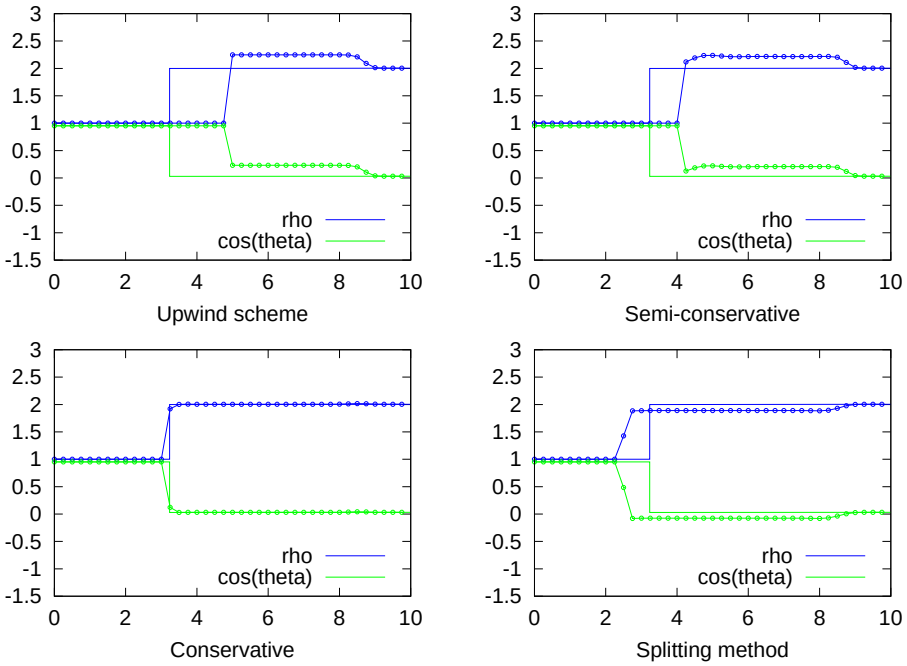


Figure III.6: Theoretical and numerical solutions of the Riemann problem (III.44)

5 The microscopic versus macroscopic Vicsek models

5.1 Local equilibrium

In this part, we would like to validate the macroscopic Vicsek model by the simulation of the microscopic Vicsek. The macroscopic model relies on the fact that the particle distribution function is at local equilibrium given by a Von Mises distribution M_Ω (see [12]):

$$M_\Omega(\omega) = C \exp\left(\frac{\omega \cdot \Omega}{d}\right) \tag{III.45}$$

where C is set by the normalization condition [1]. The goal of this section is to show numerically that the particle distribution of the microscopic Vicsek model is close in certain regimes to this Von Mises distribution.

To this aim, in appendix C we propose a numerical scheme to solve system (III.2). The setting for our particle simulations is as follows: we consider a square box with periodic boundary conditions. As initial condition for the position x_i, we choose a uniform random distribution in space. The velocity is initially distributed according to a uniform distribution on the unit circle.

During the simulation, we compute the empirical distribution of the velocity direction θ and of the mean velocity Ω of particles. We then compare this empirical distribution with its theoretical distribution $M_\Omega(\theta)$ given by (III.45). In figure III.7, we give an example of a comparison between the distribution of the velocity direction θ and the theoretical distribution M_Ω predicted by the theory.

Since the distribution of velocity converges, we have a theoretical value for the mean velocity. We denote by φ_N the mean velocity of particles and φ the theoretical value given by the stationary distribution:

$$\varphi_N = \frac{1}{N}\left|\sum_{k=1}^{N} \omega_k\right| \quad , \quad \varphi = \left|\int_\omega \omega\, M_\Omega(\omega)\, d\omega\right|. \tag{III.46}$$

At least locally in x, we have that $\varphi_N \xrightarrow{\epsilon \to 0} \varphi$. In figure III.8, we compare the two distributions for different values of the noise d and we can see that the two distributions are in good agreement. We also observe a smooth transition from order ($\varphi \approx 1$) to disorder ($\varphi << 1$) as it has been measured in the original Vicsek model [26].

The situation is different when we look at a larger system. We still have convergence of the velocity distribution of particles to a local equilibrium $\rho(x)\, M_{\Omega(x)}(\omega)$, but the mean direction $\Omega(x)$ now depends on x. Therefore the mean velocity of the particles in all the domain differs from the expected theoretical value (III.45)-(III.46). We illustrate this phenomena in figure III.9: we fix the density of particles and we increase the size of the box. As we can observe, the mean velocity φ_N (III.46) has a smaller value when the size of the box increases. This phenomena has been previously observed in [6]. The mean velocity φ_N can also differ from the expected theoretical value φ (III.45)-(III.46) when the density of particles is low. In figure III.10, we fix the size of the box ($L = 10$) and we

1. explicitly given by $C^{-1} = 2\pi\, I_0(d^{-1})$ where I_0 is the modified Bessel function of order 0

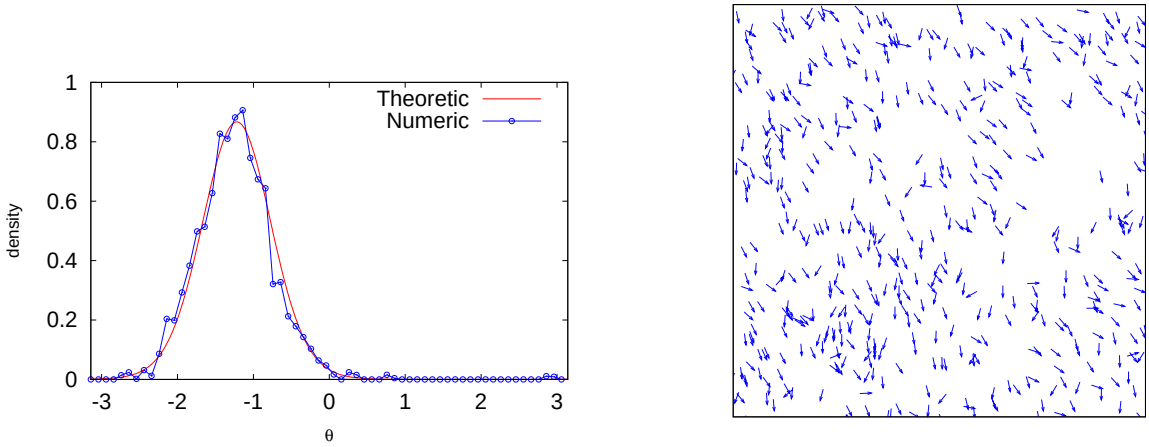


Figure III.7: Left figure: the distribution of velocity direction θ (with $\omega = (\cos\theta, \sin\theta)$) compared with its theoretical distribution after 6 time units of simulation. Right figure: the corresponding particle simulation. Parameters of the simulation: $Lx = 1$, $Ly = 1$ (domain size), number of particles $N = 500$, $\varepsilon = 1/4$, $R = .5$, $d = .2$, $\Delta t = 2.10^{-3}$.

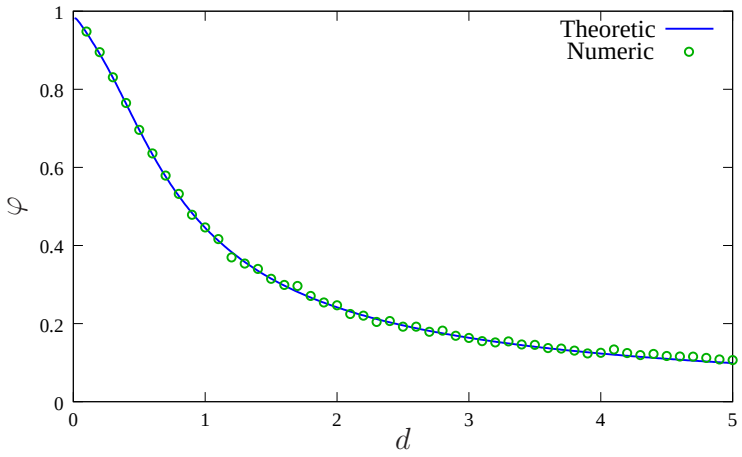


Figure III.8: The mean velocity φ (III.46) for different values of d. Parameters of the simulation: $Lx = 1$, $Ly = 1$ (domain size), number of particles $N = 200$, radius of interaction $R = .5$, $\Delta t = .02$ unit time, the simulations are run during 180 unit time.

increase the density of particles (the density is given by the number of particles inside the circle of interaction). At low density, the mean velocity φ_N is much more smaller than the theoretical prediction φ. But as the density of particles increases, the mean velocity φ_N grows (see also [26]) and moreover φ_N converges to φ. Because of that, a dense regime of particles has to be used in the following in order to numerically compare the microscopic model with the MV model.

5.2 Microscopic versus Macroscopic dynamics

We now compare the evolution of the two macroscopic quantities ρ and Ω for the two models. We have seen that the different schemes applied to the macroscopic equation can give different solutions (see figure III.6). Therefore, we expect that particle simulations will indicate what is the physically relevant solution of the macroscopic equation.

We first briefly explain how we proceed to run the particle simulations of a Riemann problem (see also appendix C). First, we have to choose a left state (ρ_l, θ_l), a right state (ρ_r, θ_r) and the noise parameter d. Then we distribute a proportion $\frac{\rho_l}{\rho_l+\rho_r}$ of particles uniformly in the interval $[0, 5]$ and the remaining particles uniformly in the interval $[5, 10]$. Then, we generate velocity distribution θ for the particles according to the distribution M_Ω (III.45) with $\Omega_l = (\cos\theta_l, \sin\theta_l)^T$ on the left side and $\Omega_r = (\cos\theta_r, \sin\theta_r)^T$ on the right side. We use the numerical scheme given in appendix C to generate particle trajectories. To make the computation simpler, we choose periodic boundary conditions. Therefore the number of particles is conserved. As a consequence, there are two Riemann problems corresponding to discontinuities at $x = 5$ and at $x = 0$ or 10 (which is the same by periodicity). We use a particle-in-cell method [14, 18] to estimate the two macroscopic quantities: the density ρ and the direction of the flux Ω (which gives θ). In order to reduce the noise due to the finite number of particles, we take a mean over several simulations to estimate the density ρ and θ (10 simulations in our examples).

In figure III.11, we show a numerical solution for the following Riemann problem:

$$(\rho_l,\, \theta_l) = (1,\, 1.5) \quad , \quad (\rho_r,\, \theta_r) = (2,\, 1.83) \quad , \quad d = 0.2 \tag{III.47}$$

using particle simulations and the macroscopic equation. We represent the solutions in a 2D representation. Since the initial condition is such that the density ρ and the direction θ are independent of the y-direction, we only represent ρ and θ along the x-axis in the following figures.

In figure III.12, we represent the two solutions (the particle and the macroscopic one) with only a dependence in the x-direction. Three quantities are represented: the density (blue), the flux direction θ (green) and the variance of the angle distribution (red). The macroscopic model supposes that the variance of θ should be constant everywhere. Nevertheless, we can see that the variance is larger in regions where the density is lower. For ρ and θ, we see clearly the propagation of a shock in the middle of the domain and a rarefaction at the boundary. The CPU time for one numerical solution at the particle level is about 140 seconds with the parameters given in figure III.12. For the macroscopic equation, the CPU time is about 0.1 second which represents a cost reduction of three orders of magnitude compared with the particle simulations. Since we have to take a mean over many particle simulations, the cost reduction is even larger.

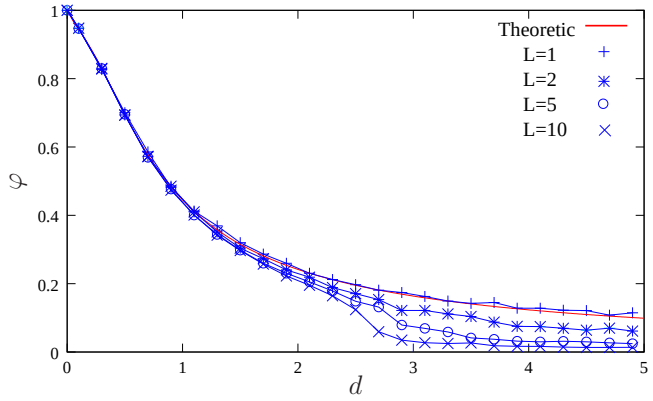


Figure III.9: The mean velocity φ (III.46) for different values of d. We use different domain sizes and we keep the same density of particles. As the domain size increases, the total flux φ decreases which means that particles are less aligned globally. Parameters of the simulations: $L = 1, 2, 5, 10$ (domain size), number of particles $N = 200, 800, 5000, 20000$, radius of interaction $R = .5$, $\Delta t = .02$ time units, the simulations are run during 180 time units.

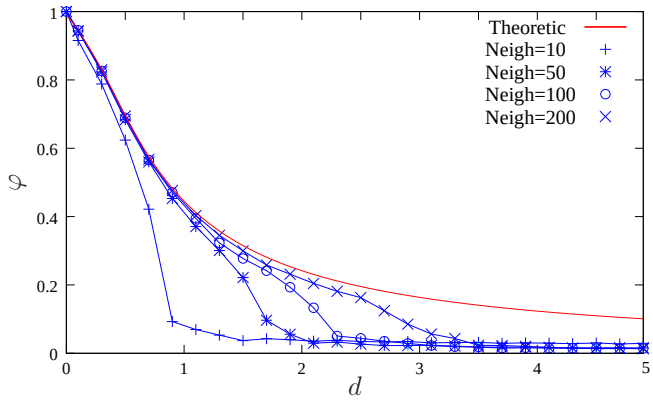


Figure III.10: The mean velocity φ (III.46) for different values of d. We change the density of particles (given by the mean number of neighbors in unit of radius of interaction). When we increase the mean number of neighbors, particles are more aligned. Parameters of the simulations: $L = 10$ (domain size), number of particles $N = 254, 1273, 6366, 12732$, and 25464, radius of interaction $R = .5$, $\Delta t = .02$ time units, the simulations are run during 180 time units.

In figure III.13, we use the same Riemann problem to set the initial position as in figure III.6 both with $d = 1$ (III.44). We use a larger domain in x ($L = 20$ space units) in order to avoid the effect of the periodic boundary condition. The upwind scheme and the semi-conservative method are clearly not in accordance with the particle simulations. Moreover, the splitting method is in better agreement with the particle simulation since the shock speed is closer to the values given by the particle simulations than that predicted by the conservative scheme.

Finally, our last simulations concern a contact discontinuity. We simply initialize with:

$$(\rho_l, \theta_l) = (1, 1) \quad , \quad (\rho_r, \theta_r) = (1, -1) \quad , \quad d = 0.2, \tag{III.48}$$

i.e. we reflect the angle with respect to the x-axis across the middle point $x = 5$. A natural solution for this problem is the contact discontinuity propagating at speed $c\,cos(1)$:

$$\rho(t, x) = 1 \quad , \quad \theta(t, x) = \theta_0(x - c\cos(1)t), \tag{III.49}$$

with $\theta_0(x) = -1$ when $x < 5$ and $\theta_0(x) = 1$ when $x > 5$. This is the solution provided by the conservative scheme (figure III.14). But surprisingly, the splitting method and the particle simulation agree on a different solution. Indeed, the solutions given by the particles and the splitting method are in fairly good agreement with each other, which seems to indicate that the "physical solution" to the contact problem (III.48) is not given by the conservative formulation (III.49) but by a much more complex profile. The constraint of unit speed drastically changes the profile of the solution compared with what would be found for a standard system of conservative laws.

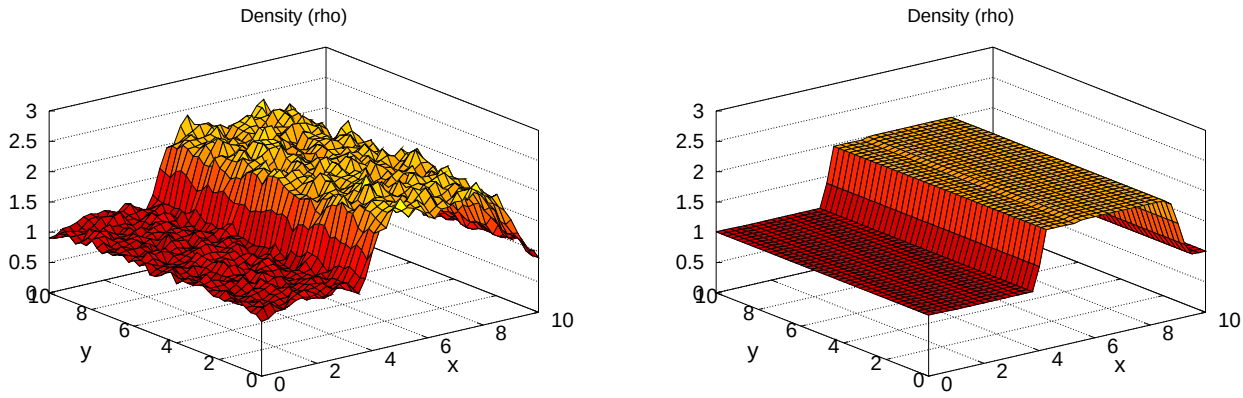


Figure III.11: The particle density in space ρ computed with particle simulations (left) and the macroscopic equations (right). We initialize with a Riemann problem (III.47). Numerical parameters for the particle simulations: $N = 2.10^6$ particles, $\Delta t = .01$, $\varepsilon = 1/10$, $R = .5$, $Lx = Ly = 10$, we take a mean over 10 computations. Numerical parameters for the macroscopic model: $\Delta t = .01$, $\Delta x = .025$ (CFL=0.416), we use the splitting method. The simulations are run during 2 time units.

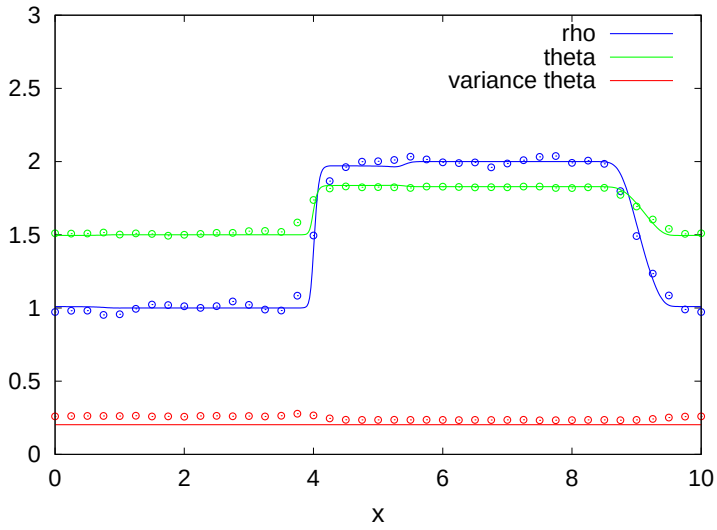


Figure III.12: The solution of the Riemann problem (III.47) with $d = .2$ computed with the splitting method (solid line) and with particle simulations (dots). In blue, we represent the density ρ, in green the flux direction θ and in red the variance of the velocity direction. The parameters are the same as in figure III.11. We only change the representation of the solution (1D-representation).

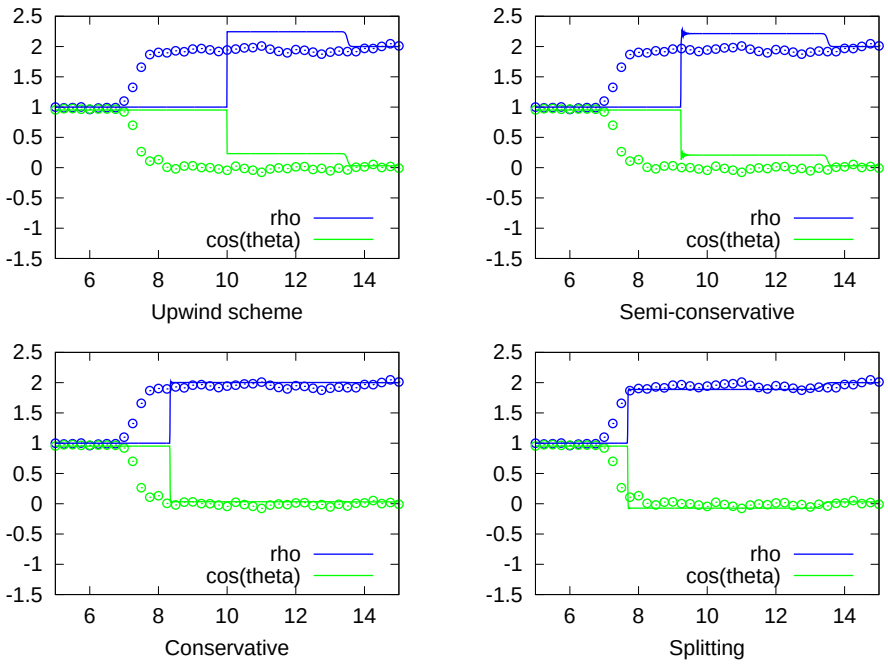


Figure III.13: The solutions of the Riemann problem (III.44) with $d = 1$ computed using the macroscopic model and particle simulations of the microscopic model (see figure III.12). Numerical parameters for the macroscopic model: $\Delta t = .01$, $\Delta x = .025$ (CFL=0.778). Numerical parameters for the particle simulation: $N = 2.10^6$ particles, $\Delta t = .02$, $\varepsilon = .1$, $R = .5$, $Lx = 20$ and $Ly = 1$. We take a mean over 50 simulations. The simulations are run during 6 time units. Since $d = 1$, fluctuations are higher (see figure III.9), we have to increase the density of particles to reduce this effect.

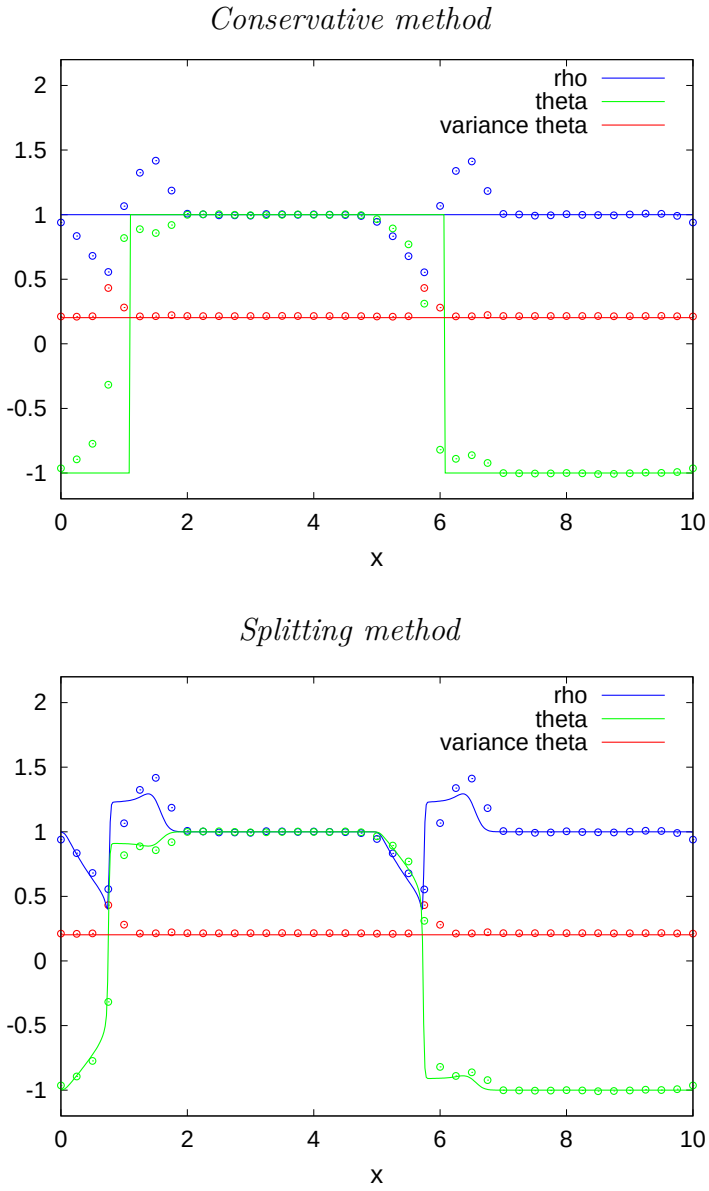


Figure III.14: The solution of the Riemann problem (III.48) computed with the conservative method (top), the splitting method (down) and with particle simulations (dots). Numerical parameters for the macroscopic model: $\Delta t = .01$, $\Delta x = .025$ (CFL=0.416). Numerical parameters for the particle simulations: $N = 10^6$ particles, $\Delta t = .01$, $\varepsilon = 1/10$, $R = .5$, $Lx = 10$, $Ly = 1$. We take a mean over 100 simulations. The simulations are run during 2 time units.

6 Conclusion

In this work, we have numerically studied both the microscopic Vicsek model and its macroscopic version [12]. Due to the geometric constraint that the velocity should be of norm one, the standard theory of hyperbolic systems is not applicable. Therefore, we have proposed several numerical schemes to solve it. By comparing the numerical simulations of the microscopic and macroscopic equations, it appears that the scheme based on a relaxation formulation of the macroscopic model, used in conjunction with a splitting method is in good agreement with particle simulations. The other schemes do not show a similar good agreement. In particular, with an initial condition given by a contact discontinuity, the splitting method and the microscopic model provide a similar solution which turn to be much more complex than what we could be expected.

These results confirm the relevance of the macroscopic Vicsek model. Since the CPU time is much lower with the macroscopic equation, the macroscopic Vicsek model is an effective tool to simulate the Vicsek dynamics in a dense regime of particles.

Many problems are still open concerning the macroscopic Vicsek model. We have seen that the splitting method gives results which are in accordance with particle simulations. But, we have to understand why this particular scheme captures well the particle dynamics better than the other schemes. Since the macroscopic equation has original characteristics, this question is challenging. Another point concerns the particle simulations. We have seen that the particles density has a strong effect on the variance of the velocity distribution. When the density is low, the variance is larger. The macroscopic equation does not capture this effect since the variance of the distribution is constant. Works in progress aims at taking into account this density effect.

A Appendix : The coefficients c_1, c_2 and λ

The analytical expression of the coefficient c_1 involved the distribution of the local equilibrium M_Ω (III.45). The two other coefficients (c_2 and λ) involve also the solution g of the following elliptic equation [12]:

$$-(1-x^2)\,\partial_x[\mathrm{e}^{x/d}(1-x^2)\partial_x g] + \mathrm{e}^{x/d} g = -(1-x^2)^{3/2}\mathrm{e}^{x/d}, \tag{III.50}$$

on the interval $x \in (-1,1)$.

If we define the function $h = \frac{g}{\sqrt{1-x^2}}$ and $M(x) = \mathrm{e}^{\frac{x}{d}}$, these macroscopic coefficients can be written as:

$$c_1 = <\cos\theta>|_M = \frac{\int_0^\pi \cos\theta\, M(\cos\theta)\sin\theta\, d\theta}{\int_0^\pi M(\cos\theta)\sin\theta\, d\theta}, \tag{III.51}$$

$$c_2 = <\cos\theta>|_{\sin^2\theta\, hM} = \frac{\int_0^\pi \cos\theta\, \sin^2\theta\, h(\cos\theta) M(\cos\theta)\sin\theta\, d\theta}{\int_0^\pi \sin^2\theta\, h(\cos\theta) M(\cos\theta)\sin\theta\, d\theta}, \tag{III.52}$$

$$\lambda = d. \tag{III.53}$$

In the above expressions, we can see that $0 \le c_1, c_2 \le 1$.

Now we are going to explore two asymptotics of g when the parameter d is small or large.

Lemma 9. *Let g be the solution of equation (III.50). We have the asymptotics:*

$$g \overset{d\to 0}{\sim} d\left[a\sin(x) - \frac{\pi}{2}\right] + o(d), \tag{III.54}$$

$$g \overset{d\to\infty}{\sim} -\frac{1}{2}\sqrt{1-x^2} + \frac{1}{12\,d}x\sqrt{1-x^2} + o\left(\frac{1}{d}\right). \tag{III.55}$$

Proof (formal). Introducing the Hilbert space:

$$V = \{g \,|\, (1-\mu^2)^{-1/2} g \in L^2(-1,1), \quad (1-\mu^2)^{1/2}\partial_\mu g \in L^2(-1,1)\}$$

we have already seen in [12] that there exists a unique solution g of (III.50). Moreover this solution is negative.

To derive the asymptotic behavior of g depending on d, we first develop (III.50):

$$\partial_x[(1-x^2)\partial_x g] + (1-x^2)\,\frac{1}{d}\partial_x g - \frac{g}{1-x^2} = (1-x^2)^{1/2}. \tag{III.56}$$

When $d \to 0$, we have:

$$\partial_x g = 0$$

on the interval $[-1+\varepsilon, 1-\varepsilon]$ for all $\varepsilon > 0$. Since g belongs to V, we also have the boundary condition $g(-1) = g(1) = 0$, so g converges to 0 when $d \to 0$.
To derive the next order of convergence in the limit $d \to 0$, we normalize g with $g = d\widetilde{g}$, which gives:

$$d\,\partial_x[(1-x^2)\partial_x \widetilde{g}] + (1-x^2)\partial_x \widetilde{g} - d\,\frac{\widetilde{g}}{1-x^2} = (1-x^2)^{1/2}. \tag{III.57}$$

In the limit $d \to 0$, we deduce that:

$$(1-x^2)\partial_x \widetilde{g} = (1-x^2)^{1/2}, \tag{III.58}$$

which has an explicit solution: $\widetilde{g} = \mathrm{asin}(x) + c$. Since $\widetilde{g} \le 0$, we have $c \le -\frac{\pi}{2}$. Numerically, we find that $c = -\frac{\pi}{2}$ but the proof is still open. This formally proves (III.54).

When $d \to +\infty$, (III.56) gives:

$$\partial_x[(1-x^2)\partial_x g_0] - \frac{g_0}{1-x^2} = (1-x^2)^{1/2}. \tag{III.59}$$

A simple calculation shows that $g_0 = -\frac{1}{2}\sqrt{1-x^2}$ is a solution of (III.59).
To derive the next order of convergence, we look at the difference $v = d(g - g_0)$, which satisfies (see (III.56) and (III.59)):

$$\partial_x[(1-x^2)\partial_x v] + (1-x^2)\,\frac{1}{d}\partial_x v - \frac{v}{1-x^2} = -(1-x^2)\,\partial_x g_0.$$

In the limit $d \to +\infty$, v satisfies:

$$\partial_x[(1-x^2)\partial_x v] - \frac{v}{1-x^2} = -\frac{1}{2}x\sqrt{1-x^2}. \tag{III.60}$$

A simple calculation shows that $v = \frac{1}{12}x\sqrt{1-x^2}$ is solution of (III.60). Therefore we formally have the expression (III.55) in the proposition.

∎

In figure III.15, we compute numerically the function g (III.50). We use a finite element method with a space step $\Delta x = 10^{-3}$. The two asymptotics of g when $d \to 0$ and $d \to +\infty$ are computed in figure III.16.

Proposition 10. *The two coefficients c_1 and c_2 defined (resp.) by the equations (III.51) and (III.52) satisfy the following asymptotics:*

$$c_1 \overset{d\to 0}{\sim} 1 - d + O(d^2), \tag{III.61}$$

$$c_1 \overset{d\to +\infty}{\sim} \frac{1}{3d} + O\left(\frac{1}{d^2}\right), \tag{III.62}$$

$$c_2 \overset{d\to \infty}{\sim} \frac{1}{6d} + o\left(\frac{1}{d}\right). \tag{III.63}$$

Proof. We have an explicit expression for the coefficient c_1 using the change of unknowns $x = \cos(\theta)$:

$$c_1 = \coth\left(\frac{1}{d}\right) - d, \tag{III.64}$$

where $\coth(s) = \frac{e^s + e^{-s}}{e^s - e^{-s}}$. The expressions of (III.61) and (III.62) are simply deduced by a Taylor expansion of the last expression.

For the coefficient c_2, we insert the development of g (III.55) in expression (III.52). ∎

Remark. The behavior of c_2 when $d \to 0$ is more difficult to analyze. The density probability $\sin\theta\, h\, M$ used in formula (III.52) becomes singular in this limit. Nevertheless, due to the expression of M_Ω, the density converges to a Dirac delta at 0 which explains why $c_2 \overset{d\to 0}{\sim} 1$. To capture the next order of convergence, we need to find the second order correction of g in the limit $d \to 0$ which is not available. However, numerically we find that:

$$c_2 \overset{d\to 0}{\sim} 1 - 2d + o(d).$$

In figure III.17, we numerically compute the coefficients c_2/c_1, λ/c_1 and their asymptotics.

B Appendix : Special solution of the MV model

In this appendix, a vortex configuration is exhibited as a stationary solution of the MV model (III.4)-(III.5)-(III.6) in dimension 2. A stationary state of the MV model has to satisfy:

$$\begin{aligned} &\nabla_x \cdot (\rho\Omega) = 0, \\ &c(\Omega \cdot \nabla_x)\Omega + \lambda\,(\mathrm{Id} - \Omega \otimes \Omega)\tfrac{\nabla_x \rho}{\rho} = 0. \end{aligned} \tag{III.65}$$

Introducing the polar coordinates, $\rho(r,\theta)$, $\Omega(r,\theta) = f_r(r,\theta)\vec{e}_r + f_\theta(r,\theta)\vec{e}_\theta$, where $\vec{e}_r = (\cos\theta, \sin\theta)^T$ and $\vec{e}_\theta = (-\sin\theta, \cos\theta)^T$, we are able to formulate the proposition:

Proposition 11. *The following initial condition is a stationary state of the MV model (III.65):*

$$\rho(r) = C\, r^{c/\lambda} \qquad , \qquad \Omega = \vec{e}_\theta, \tag{III.66}$$

where C is a constant.

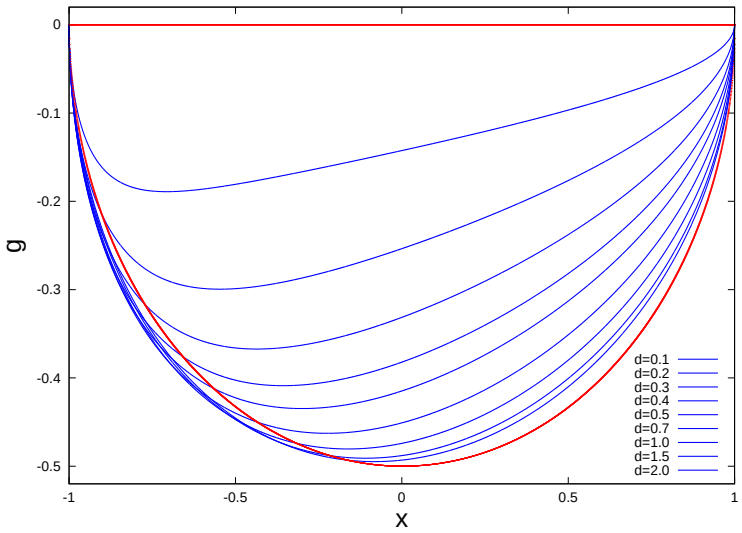


Figure III.15: The numerical solution g of (III.50) for different values of the parameter d. We have the following asymptotics (see lemma 9): $g \overset{d\to 0}{\longrightarrow} 0$ and $g \overset{d\to\infty}{\longrightarrow} -\frac{1}{2}\sqrt{1-x^2}$.

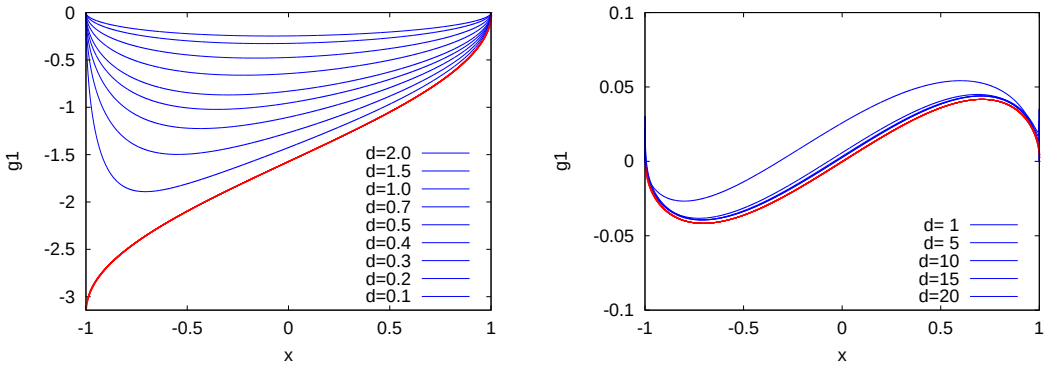


Figure III.16: Left figure: the first correction g_1 of g when $d \to 0$. The red curve is the theoretical asymptotic limit: $g_1 = g/d \overset{d\to 0}{\longrightarrow} \operatorname{asin}(x) - \frac{\pi}{2}$ (see lemma 9). Right figure: the first correction g_1 of g when $d \to +\infty$. The red curve is the theoretical asymptotic: $g_1 = d(g + \frac{1}{2}\sqrt{1-x^2}) \overset{d\to 0}{\longrightarrow} \frac{1}{12}x\sqrt{1-x^2}$ (see lemma 9).

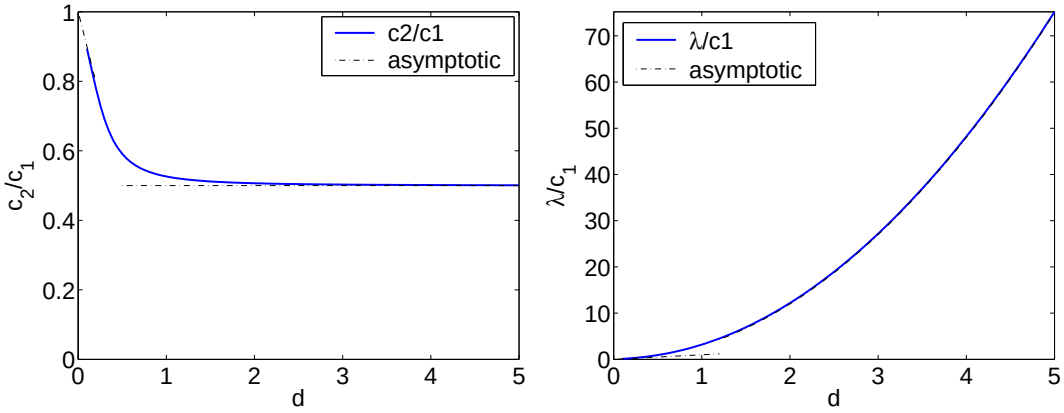


Figure III.17: The ratio c_2/c_1 and λ/c_1 (solid lines) and their two asymptotics (dashed lines) (see proposition 10) computed with $\Delta x = 10^{-3}$.

Proof. With the expression of ρ and Ω given by (III.66), the divergence of the mass is zero and the gradient of ρ is orthogonal to Ω, therefore the system (III.65) reduces to:

$$c(\Omega \cdot \nabla_x)\Omega + \lambda \frac{\nabla_x \rho}{\rho} = 0, \tag{III.67}$$

or in polar coordinates:

$$c\frac{1}{r}\partial_\theta \vec{e}_\theta \; + \; \lambda \frac{\rho'(r)}{\rho(r)}\vec{e}_r = 0.$$

Since $\partial_\theta \vec{e}_\theta = -\vec{e}_r$, we can easily check that the solution of this equation is given by $\rho(r) = C\, r^{c/\lambda}$. ∎

C Appendix: Numerical schemes for particle simulations

In the limit $\varepsilon \to 0$, an explicit Euler method for the differential system (III.1)-(III.2) imposes a restriction time step condition of $\frac{1}{\varepsilon}\Delta t < 1$. Therefore, we develop an implicit scheme for this system. The idea is to go back to the original Vicsek model (see [12]). We use the formulation:

$$\frac{\omega^{n+1} - \omega^n}{\Delta t} = (\mathrm{Id} - \omega^{n+1/2} \otimes \omega^{n+1/2})(\bar{\omega}^n - \omega^n) \tag{III.68}$$

where $\omega^{n+1/2} = \frac{\omega^n + \omega^{n+1}}{|\omega^n + \omega^{n+1}|}$ and $\bar{\omega}^n$ is the average velocity (III.3). When $\Delta t = 1$, we recover exactly the original Vicsek model [26]. (III.68) can in fact be solved explicitly. First, we have to remember that ω^{n+1} belongs to the unit circle (i.e. $|\omega^{n+1}| = 1$). Then we use that $\omega^{n+1} - \omega^n$ is the orthogonal projection of $(\bar{\omega}^n - \omega^n)\Delta t$ on the orthogonal plan of $\omega^{n+1/2}$. Therefore ω^{n+1} and ω_n are on the circle $\mathcal{C}$ with center $B = \omega_n + \frac{(\bar{\omega}^n - \omega^n)\Delta t}{2}$ and radius $\left|\frac{(\bar{\omega}^n - \omega^n)\Delta t}{2}\right|$ (see figure III.18). This fully defines ω^{n+1} since ω^n and ω^{n+1} are the two intersection points of the unit circle and the circle $\mathcal{C}$. Denoting θ the angle of the

unit vector ω, we easily check that we have in terms of angles:

$$\theta^{n+1} = \theta^n + 2(\widehat{\omega^n, B}).$$

To take into account the effect of the noise, we simply add a random variable:

$$\theta^{n+1} = \theta^n + 2(\widehat{\omega^n, B}) + \sqrt{2d\,\Delta t}\,\epsilon_n \tag{III.69}$$

where ϵ_n is a standard normal distribution independent of θ^n.

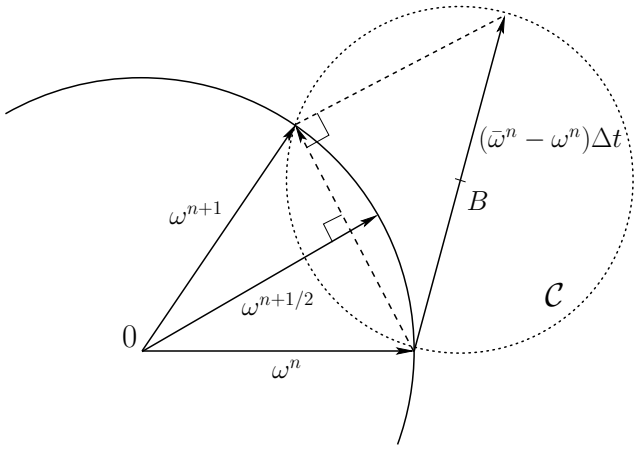


Figure III.18: Illustration of the geometric method to solve explicitly equation (III.68).

Algorithm used to solve a Riemann problem with particles.

1. Choose a Riemann problem $(\rho_l,\, \theta_l)$ and $(\rho_r,\, \theta_r)$.
2. Initiate N particles $(x_k, \omega_k)_{k=1..N}$ according to the distributions $\rho_l M_{\Omega_l}$ and $\rho_r M_{\Omega_r}$.
3. Let evolve the particles in time using the time-discretization (III.69) of equation (III.2).
4. Compute the mass ρ and the direction of the flux Ω using Particle-In-Cell method [18] in order to compare the simulation with the one of the MV model.

Bibliography

[1] I. Aoki. A simulation study on the schooling mechanism in fish. *Bulletin of the Japanese Society of Scientific Fisheries*, 48(8):1081–1088, 1982.

[2] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, 105(4):1232, 2008.

[3] E. Bertin, M. Droz, and G. Gregoire. Boltzmann and hydrodynamic description for self-propelled particles. *Phys. Rev. E*, 74(2):22101, 2006.

[4] F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*. Birkhauser, 2004.

[5] J. Buhl, D.J.T. Sumpter, I.D. Couzin, J.J. Hale, E. Despland, E.R. Miller, and S.J. Simpson. From disorder to order in marching locusts, 2006.

[6] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud. Collective motion of self-propelled particles interacting without cohesion. *Phy. Rev. E*, 77(4):46113, 2008.

[7] G.Q. Chen, C.D. Levermore, and T.P. Liu. Hyperbolic conservation laws with sti relaxation and entropy. *Comm. Pure Appl. Math*, 47:787, 1994.

[8] Y. Chuang, M.R. D'Orsogna, D. Marthaler, A.L. Bertozzi, and L.S. Chayes. State transitions and the continuum limit for a 2d interacting, self-propelled particle system. *Physica D*, 232(1):33–47, 2007.

[9] I.D. Couzin, J. Krause, R. James, G.D. Ruxton, and N.R. Franks. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218(1):1–11, 2002.

[10] F. Cucker and E. Mordecki. Flocking in noisy environments. *Journal de mathématiques pures et appliquées*, 2007.

[11] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862, 2007.

[12] P. Degond and S. Motsch. Macroscopic limit of self-driven particles with orientation interaction. *C. R. Acad. Sci. Paris, Ser.I*, 345(10):555–560, 2007.

[13] P. Degond and S. Motsch. Large scale dynamics of the persistent turning walker model of fish behavior. *J. Stat. Phys.*, 131(6):989–1021, 2008.

[14] H. Fehske, R. Schneider, and A. Weisse. *Computational Many-Particle Physics*. Springer Verlag, 2007.

[15] J. Gautrais, C. Jost, M. Soria, A. Campo, S. Motsch, R. Fournier, S. Blanco, and G. Theraulaz. Analyzing fish movement as a persistent turning walker. *J. Math. Biol.*, 58(3):429–445, 2009.

[16] G. Grégoire and H. Chaté. Onset of collective and cohesive motion. *Phys. Rev. Lett.*, 92(2):25702, 2004.

[17] S.Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic descriptions of flocking. *Kinet. Relat. Models*, 1(3):415–435, 2008.

[18] R.W. Hockney and J.W. Eastwood. *Computer Simulation Using Particles*. Institute of Physics Publishing, 1988.

[19] A. Huth and C. Wissel. The simulation of the movement of fish schools. *J. Theor. Biol.*, 156(3):365–385, 1992.

[20] P.L. LeFloch. Entropy weak solutions to nonlinear hyperbolic systems under non-conservative form. *Comm. in PDE*, 13(6):669–727, 1988.

[21] R.J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, 1992.

[22] R.J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge University Press, 2002.

[23] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *J. Math. Biol.*, 38(6):534–570, 1999.

[24] M. Nagy, I. Daruka, and T. Vicsek. New aspects of the continuous phase transition in the scalar noise model (SNM) of collective motion. *Physica A*, 373:445–454, 2007.

[25] C.W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34, 1987.

[26] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75(6):1226–1229, 1995.

Troisième partie : Limite de congestion

Chapitre IV

La congestion dans un modèle macroscopique pour animaux grégaires

Ce chapitre a été écrit en collaboration avec Pierre Degond, David Sanchez et Richard Bon. Il a fait l'objet de la publication suivante :

P. Degond, L. Navoret, R. Bon, D. Sanchez, ***Congestion in a macroscopic model of self-driven particles modeling gregariousness****, J. Stat. Phys., 138 :85-125, 2010.*

Nous avons ajoutés deux courtes appendices : la première précisant la nature du mouvement dans les domaines congestionnés et la deuxième présentant des simulations numériques du système particulaire. La suite de ce chapitre est écrite en anglais.

Congestion in a macroscopic model of self-driven particles modeling gregariousness

Abstract: We analyze a macroscopic model with a maximal density constraint which describes short range repulsion in biological systems. This system aims at modeling finite-size particles which cannot overlap and repel each other when they are too close. The parts of the fluid where the maximal density is reached behave like incompressible fluids while lower density regions are compressible. This paper investigates the transition between the compressible and incompressible regions. To capture this transition, we study a one-dimensional Riemann problem and introduce a perturbation problem which regularizes the compressible-incompressible transition. Specific difficulties related to the non-conservativity of the problem are discussed.

1 Introduction

We consider a macroscopic model of self-driven particles which describes the dynamics of a large number of social interactive agents. More specifically, we are interested in modeling short range repulsion effects due to the fact that finite-size agents (e.g. sheep in a herd) cannot overlap (non-overlapping or steric constraints). To this aim, we derive a hyperbolic problem with a density constraint as a limit of an unconstrained system with a repulsive force which turns on suddenly when the density becomes close to the

maximal one. The limit model requires transmission conditions at the transition between an unclustered region (where the maximal density is not reached) and a clustered region. In unclustered regions, the fluid is compressible while it becomes incompressible in the clustered ones. Therefore, this paper aims at providing a description of this transition between a compressible and an incompressible fluid. Unfortunately, the formal perturbative approach which we implement does not directly provide information about these transmission conditions. In order to retrieve this information, we rigorously analyze special solutions of the perturbation problem: the Riemann problem. These solutions are explicitely known and allow us to carry out the limit rigorously and to recover the required transmission conditions. We postulate that these conditions, which are rigorously proven only for Riemann problem solutions, do extend to all solutions. However, being non-rigorous for general solutions, these conditions are stated as formal conditions in the "formal statement 1" below, which constitutes the main result of the present paper. Still, the rigorous analysis of Riemann problem solutions is quite technical and the proofs of many statements are deferred to appendices.

The modeling of biological systems undergoing flocking or herding dynamics has been the subject of a vast literature. A first class of models relies on the alignement interaction between neighbouring self-propelled particles. The simplest of these models is an individual-based (or microscopic) model proposed by Vicsek [26, 47]. A macroscopic version of the Vicsek model is derived in [22] and a collisional Vicsek model is proposed in [8]. A variant of the Vicsek model has been proposed by Cucker and Smale [17, 18] (see also [12, 27] for recent mathematical results). By incorporating long-range attractive and short-range repulsive forces to the Vicsek model, one obtains the three zones model of Aoki [1, 16, 42], originally devised to describe fish schools. Models with repulsive-attractive interaction only (without alignement interaction) have been studied in [14, 19, 35]. Such models have been used for pedestrian interactions [29, 37]. Other kinds of macroscopic models of drift-diffusion type have been analyzed in [28, 34, 44, 46] and different hyperbolic models are compared in [5]. For biological reviews, we can refer to [15, 36].

As outlined above, we focus here on the congestion constraint: animals or individuals cannot overlap (steric constraint). As a consequence this congestion constraint leads to the existence of a maximal density ρ^*, which cannot be exceeded inside the flock. This problem has been analyzed before and schematically two methods have been proposed. A first one consists in modeling repulsion through forces or diffusion terms [14, 19, 34, 35, 46]. However, in this approach, the individuals are point particles and their finite size is not explicitely described. So the maximal density constraint is not explicitely taken into account. To explicitely take this maximal density constraint inco account, in [32, 33], the authors have developped an alternative approach: the particles are first evolved freely over one time step and then projected towards the "closest" admissible non-overlapping configuration. This leads to non-local interaction between the particles which contradicts the local character of the interactions in most biological systems. By contrast, we developped a third route inspired by multi-phase flows [10] and traffic jam modeling [6, 7]. The repulsive force is modeled by a nonlinear pressure law $p(\rho)$ which becomes singular as the density approaches the maximal density ρ^*. Additionnally a small parameter ε allows to describe the fact that the regularized pressure is very small of order ε as long as the density ρ is smaller than ρ^* and turns on suddenly to a finite

or even large value when ρ becomes close to ρ^*. In the limit $\varepsilon \to 0$ of this model, two distinct phases appear: a pressureless compressible phase which describes free motion in unclustered regions and an incompressible phase which describes the motion inside the clusters. The major difficulty is to find the transmission conditions between the compressible and incompressible phases.

The present paper is a multi-dimensional extension of the methodology presented in [6, 7, 10] for multi-phase flows or traffic. However, an additional difficulty arises due to the non-conservative character of the original hyperbolic model. Indeed, momentum is not a conserved quantity because the particles in the underlying particle system are self-propelled particles which have constant (in-time) and uniform (in-space) velocities. Therefore, the model which is at the starting point of this paper is a non-conservative hyperbolic system which as such presents an ambiguity in the definition of weak solutions. We will show that this ambiguity can be partly removed for one-dimensional Riemann problem solutions. We believe that the strategy developped in this paper to analyze congestion effects can apply to other systems such as bacteria populations [38], economic systems like supply chains [2] or physical systems like granular materials [4, 41].

The organization of this article is as follows. In section 2, we present the perturbation model and its limit. We also provide the connection conditions between the compressible and incompressible phases of the limit model, which are the main result of the paper. A remark on collision of clusters is also formulated. With these informations, we show that the available information is sufficient to provide a well-defined dynamics at least in the case of a single cluster. Section 3 is devoted to the study of the one-dimensional Riemann problem for the perturbed problem and the limits of its solutions as $\varepsilon \to 0$. As stated above, this analysis provides a strong support for (but not a proof of) the postulated transmission conditions at the compressible-incompressible interface which are provided in section 2. Appendix A provides a formal derivation of the initial model from an individual based model with long-range attractive and short-range repulsive interactions, which describes the aggregation of gregarious animals like sheep. Appendices B to E provide proofs of technical lemmas and theorems needed in the analysis of the Riemann problem.

2 Model and goals

2.1 The model and its rescaled form

Our starting point is the following model, written in dimensionless form:

$$\partial_t \rho + \nabla_x \cdot (\rho \Omega) = 0, \tag{IV.1}$$
$$\partial_t \Omega + (\Omega \cdot \nabla_x)\Omega + (\mathrm{Id} - \Omega \otimes \Omega)\nabla p(\rho) = 0, \tag{IV.2}$$

where $\rho = \rho(x,t)$ is the particle density and $\Omega = \Omega(x,t)$ is the particle velocity. The problem is posed on the 2-dimensional plane $x \in \mathbb{R}^2$ and $t > 0$ is the time. The velocity $\Omega(x,t) \in \mathbb{R}^2$ is supposed to satisfy the normalization constraint

$$|\Omega(x,t)| = 1, \ \forall x \in \mathbb{R}^2, \ \forall t > 0.$$

Therefore, $\Omega(x,t) \in \mathbb{S}^1$, the unit sphere, at any point in space-time. The function $p(\rho)$ is an increasing function such that $p(\rho) \sim \rho^\gamma$ when $\rho \ll 1$ and $p(\rho) \to +\infty$ when $\rho \to \rho^*$ where ρ^* is the so-called congestion density. In this paper, we will consider

$$p(\rho) = \frac{1}{\left(\frac{1}{\rho} - \frac{1}{\rho^*}\right)^\gamma}, \tag{IV.3}$$

for simplicity but any other function with similar behaviour would lead to similar results. Note that a convenient choice of the density scale allows us to make the constant at the numerator of (IV.3) equal to one. The operators $\nabla_x \cdot$ and $(\Omega \cdot \nabla_x)$ are defined, for a vector field $A = (A_1, A_2)(x)$, by

$$\begin{aligned}
&\nabla_x \cdot A = \partial_{x_1} A_1 + \partial_{x_2} A_2, \\
&(\Omega \cdot \nabla_x) A = ((\Omega_1 \partial_{x_1} + \Omega_2 \partial_{x_1}) A_1, (\Omega_1 \partial_{x_1} + \Omega_2 \partial_{x_1}) A_2)^T,
\end{aligned}$$

where T denotes the transpose operator. Finally, $(\mathrm{Id} - \Omega \otimes \Omega)$ is the projection matrix onto the line spanned by $\Omega^\perp$, where $\Omega^\perp$ is the vector Ω rotated by the angle $\pi/2$. Alternatively, we have, for a vector A:

$$(\mathrm{Id} - \Omega \otimes \Omega) A = A - (\Omega \cdot A) A,$$

where $(\Omega \cdot A)$ is the dot product $\Omega \cdot A = \Omega_1 A_1 + \Omega_2 A_2$.

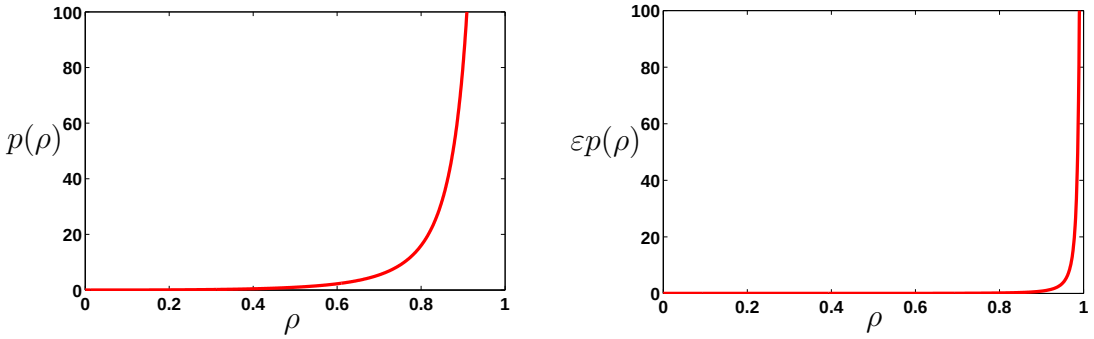
We show in appendix A that this model well describes the behaviour of a system of particles subjected to long-range attraction and short-range repulsion in the spirit of a model proposed by Aoki [1] or Couzin et al [16] for modelling gregariousness and swarming. More precisely, in appendix A, we derive this system from such a particle system through successive changes of scales via mean-field and hydrodynamic theories. In the form (IV.2), we have dropped the force term describing long-range attraction. Indeed, this force term would add the quantity $(\mathrm{Id} - \Omega \otimes \Omega)\xi_a$ at the right-hand side of (IV.2), with

$$\xi_a(x,t) = \frac{\int K_a(|y-x|)(y-x)\rho(y,t)dy}{\int K_a(|y-x|)\rho(y,t)dy},$$

where K_a is a bounded positive kernel. This terms does not add any differential operator and all the subsequent analysis will stay unaltered by adding this term.

Our main concern is the study of the congestion effects brought by the singularity of $p(\rho)$ near the congestion density ρ^*. Indeed, a herd of animals can be viewed, at large scales, as a domain of space where the density ρ is close to the saturation density ρ^*. Therefore, the geometrical domain occupied by the herd at time t can be identified to a set $H_t = \{x \in \mathbb{R}^2 \,|\, \rho^* - \delta\rho < \rho(x,t) < \rho^*\}$ where the parameter $\delta\rho > 0$ must be suitably tuned. Therefore, with the initial model (IV.1), (IV.2), the definition of a herd depends on an arbitrary parameter $\delta\rho$, which makes it ambiguous.

A way to unambiguously define the herd is to force the system (IV.1)-(IV.2) to make clear-cut phase transitions from unclustered $\rho < \rho^*$ to clustered $\rho = \rho^*$ phases. In the spirit of the works [6, 7, 21] for traffic, this can be achieved in an asymptotic regime which amounts to supposing that there is merely no repulsive interactions at all as long as $\rho < \rho^*$, and that repulsive "pressure" forces turn on suddenly when ρ hits the congestion density ρ^*. This can be done by rescaling $p(\rho)$ into $\varepsilon p(\rho)$ where $\varepsilon \ll 1$ is a small parameter. In



(a) $\varepsilon = 1$ (b) $\varepsilon = 10^{-2}$

Figure IV.1: The "potential" for repulsive interaction $p(\rho)$ (left) and $\varepsilon p(\rho)$ (right) after scaling by a small parameter $\varepsilon = 10^{-2}$, with $\gamma = 2$ and $\rho^* = 1$. From the right picture, it is clear that the repulsive interaction turns on only when ρ is very close to ρ^*

this way, repulsive interactions are $O(\varepsilon)$ as long as $\rho < \rho^*$, but become $O(1)$ when $\rho = \rho^*$ (see fig. **??**).

Biologically, this assumption amounts to saying that the animals do not change their directed motion by the presence of their neighbours unless they touch them and need to modify their trajectory to bypass them. The parameter $\varepsilon \ll 1$ is related to the time scale at which this change of trajectory occurs and is therefore supposed small. Let us also note that our model considers that all animals move with speed unity and never stop. Obviously the model will require improvements by taking into account the fact that a certain fraction of animals are steady, while foraging or resting.

Therefore, our main concern in this paper is the study of the following perturbation problem:

$$\partial_t \rho^\varepsilon + \nabla_x \cdot (\rho^\varepsilon \Omega^\varepsilon) = 0, \tag{IV.4}$$

$$\partial_t \Omega^\varepsilon + (\Omega^\varepsilon \cdot \nabla_x)\Omega^\varepsilon + \varepsilon(\mathrm{Id} - \Omega^\varepsilon \otimes \Omega^\varepsilon)\nabla_x p(\rho^\varepsilon) = 0, \tag{IV.5}$$

$$|\Omega^\varepsilon| = 1. \tag{IV.6}$$

We will be interested in the formal limit $\varepsilon \to 0$. A rigorous theory of this type of problems is unfortunately still out of reach up to our knowledge. In the following section, we show that the limit $\varepsilon \to 0$ leads to a phase transition between compressible and incompressible regimes.

2.2 The singular limit $\varepsilon \to 0$: transition between compressible and incompressible motion

As $\varepsilon \to 0$, $\varepsilon p(\rho^\varepsilon)$ becomes significant only where the convergence $\rho^\varepsilon \to \rho^*$ is fast enough. Therefore, in the limit, either $\rho^\varepsilon \to \rho < \rho^*$ and $\varepsilon p(\rho^\varepsilon) \to 0$ or $\rho^\varepsilon \to \rho^*$ and $\varepsilon p(\rho^\varepsilon) \to \bar{p}$ with $\bar{p}$ possibly non zero. In other words, the equation $(\rho^* - \rho)\bar{p} = 0$ holds in the limit. If additionally $\bar{p} < +\infty$, straighforward inspection shows that

$$\rho^* - \rho^\varepsilon = O(\varepsilon^{\frac{1}{\gamma}}). \tag{IV.7}$$

Therefore, the formal limit $\varepsilon \to 0$ of system (IV.4)-(IV.5)-(IV.6) is given by the following system:

$$\partial_t \rho + \nabla_x \cdot (\rho \Omega) = 0, \tag{IV.8}$$
$$\partial_t \Omega + \Omega \cdot \nabla_x \Omega + (\mathrm{Id} - \Omega \otimes \Omega) \nabla_x \bar{p} = 0, \tag{IV.9}$$
$$|\Omega| = 1, \tag{IV.10}$$
$$(\rho^* - \rho)\bar{p} = 0. \tag{IV.11}$$

In the non-congested domain $\rho < \rho^*$, the system reduces to a pressureless compressible gaz dynamics model with a speed constraint

$$\partial_t \rho + \nabla_x \cdot \rho \Omega = 0,$$
$$\partial_t \Omega + \Omega \cdot \nabla_x \Omega = 0,$$
$$|\Omega| = 1.$$

This system describes the behaviour of the system outside the congested region. It is a compressible system. Biologically, it describes the behaviour of dispersed animals outside the herd. Mathematical studies of this system are outside the scope of this article and the reader can refer to [9] for standard pressureless gas dynamics models (without speed constraint). We note that this system exhibits vacuum regions where $\rho = 0$ as it will be seen below.

2.3 Study of the congested region

The congested part of the flow is defined as the region where the congestion constraint $\rho = \rho^*$ is reached. Biologically, it defines the domain of space occupied by the herd. Its connected components will be called "clusters". In the congested domain, system (IV.8)-(IV.11) turns into an incompressible Euler model with speed constraint:

$$\nabla_x \cdot \Omega = 0, \tag{IV.12}$$
$$\partial_t \Omega + \Omega \cdot \nabla_x \Omega + (\mathrm{Id} - \Omega \otimes \Omega) \nabla_x \bar{p} = 0, \tag{IV.13}$$
$$|\Omega| = 1, \tag{IV.14}$$
$$\rho = \rho^*. \tag{IV.15}$$

We first note that smooth incompressible vector fields of constant norm in $\mathbb{R}^2$ have a very special structure which is outlined in the following.

Proposition 12. [1] *Let $\Omega(x)$ be a smooth vector field on a domain $\Theta \subseteq \mathbb{R}^2$ with values in $\mathbb{S}^1$ and which satisfies the incompressibility constraint $\nabla_x \cdot \Omega = 0$. Then the integral lines of $\Omega^\perp$ are straight lines and Ω is constant along these lines (where $\Omega^\perp$ is rotated by an angle of $\pi/2$) and the integral lines of Ω are parallel curves to each other.*

The proof of this proposition simply results from introducing the angle θ so that $\Omega(x,t) = (\cos(\theta(x,t)), \sin(\theta(x,t)))$ and noting that θ satisfies the "transport equation"

$$\partial_{x_2} \theta - (\tan \theta) \partial_{x_1} \theta = 0.$$

1. This proposition can be improved: in appendix H, we show that a smooth incompressible vector field with value in $\mathbb{S}^1$ is constant.

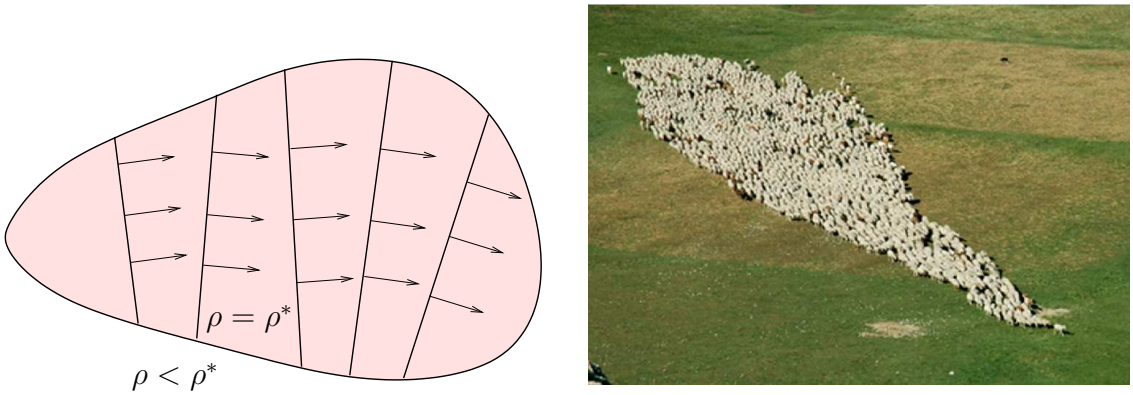


Figure IV.2: Left: schematic figure of a congested zone, where the arrows design the vectors Ω. Right: picture of a sheep herd (© Guy Theraulaz, CNRS, CRCA, Toulouse, France).

This property implies that the knowledge of Ω on the cluster boundaries suffices to know Ω everywhere inside the clusters.

The integral curves of Ω provide a mathematical description of the animal files in the herd. These curves being parallel to each other, they are consistant with the intuition and the observation of animal files in a herd (see fig. IV.2).

The pressure $\bar{p}$ satisfies an elliptic equation. Indeed, by taking the divergence of the equation (IV.13) and after easy computations, we get

$$\nabla_x \cdot ((\mathrm{Id} - \Omega \otimes \Omega)\, \nabla_x \bar{p}) = \mathrm{Tr}((\nabla_x \Omega)(\nabla_x \Omega)^T), \tag{IV.16}$$

where Tr is the trace of a matrix and the exponent T denotes the transpose operator. This equation can be equivalently written:

$$-(\Omega^\perp \cdot \nabla_x)^2 \bar{p} - (\nabla_x \cdot \Omega^\perp)(\Omega^\perp \cdot \nabla_x)\bar{p} = -\mathrm{Tr}((\nabla_x \Omega)(\nabla_x \Omega)^T), \tag{IV.17}$$

and only involves the operator $(\Omega^\perp \cdot \nabla_x)$ applied to $\bar{p}$. Since the integral lines of $\Omega^\perp$ are straight lines, equation (IV.17) is just a one-dimensional elliptic problem for $\bar{p}$ posed on this straight line. Knowing the boundary values of $\bar{p}$ where this straight line meets the boundary of the cluster allows to compute $\bar{p}$ everywhere on this lines and consequently inside the cluster (see fig. IV.2). Hence, once Ω is known inside the cluster, the resolution of this equation only requires the knowledge of the boundary conditions for $\bar{p}$ at the boundaries of the cluster.

To close the system, i.e. to determine how the solution in the congested domain evolves, we need to determine these boundary conditions. They are not given by the formal limit and, in order to determine them, we need to explore another route. For this pupose we look at the solutions of the Riemann problem for the perturbed and limit systems. Note that if we abandon the constraint of constant norm $|\Omega| = 1$, the non conservative term $(\Omega \otimes \Omega)\nabla_x \bar{p}$ in the momentum conservation equation (IV.9) drops out, and we recover a conservative model expressing mass and momentum equation. Then, the Rankine-Hugoniot conditions across the boundary between the compressible and incompressible regions provide the boundary conditions for the pressure at the cluster

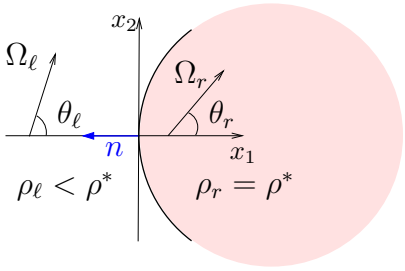


Figure IV.3: Notations at the interface.

boundary. The constant norm constraint prevents from using this strategy. Therefore, we need to find a different route to specify these boundary conditions.

2.4 Conditions at the boundary of the clusters

To find the boundary conditions on the cluster boundaries, we need to extract more information from the perturbation system (IV.4)-(IV.6) than the mere limit system (IV.8)-(IV.11). As such, this system is underdetermined. The strategy is to extract such information by passing to the limit $\varepsilon \to 0$ in some special solutions of this system. To underline the difficulty resulting from the non-conservativity, let us first look at the Rankine-Hugoniot conditions. We have the following proposition.

Proposition 13. *1. If ρ and Ω are smooth on both sides of a dicontinuity curve $\Gamma(t)$ parametrically defined by $x = x_c(s,t)$ where $s \in \mathbb{R}$ is the curvilinear abscissa along $\Gamma(t)$, then we have*

$$[\rho(\Omega \cdot n - \sigma)]_\Gamma = 0,$$

where $n = (\partial x_c/\partial s)^\perp$ is the unit normal to $\Gamma(t)$ at $x_c(s,t)$ and $\sigma = (\partial x_c/\partial t) \cdot n$ is the geometric speed of the discontinuity $\Gamma(t)$ at $x_c(s,t)$.

2. If Ω is smooth (i.e. $\mathcal{C}^1$) across Γ and ρ is smooth on both sides of Γ, then we have

$$[\bar{p}]_\Gamma (\Omega \cdot \tau) = 0,$$

where $\tau = \partial x_c/\partial s$ is a unit tangent vector to $\Gamma(t)$.

The proof of this proposition is omitted. The second relation provides us information when the mean velocity is not tangent to the cluster. In this condition, if the mean velocity is continuous, the pressure is also continuous. This implies that the pressure is zero on a cluster boundary if the mean velocity is continuous. This fact will be supported by the forthcoming analysis. However, as regards the interface dynamics, such an analysis is incomplete because the second equation supposes that Ω is continuous.

So as to capture the correct boundary conditions for the pressure $\bar{p}$ and the velocity Ω at a cluster boundary where ρ and Ω may be discontinuous, we consider a one dimensional problem in the normal direction n to this boundary (cf. figure IV.3). In order to justify this simplication, we introduce the coordinate system (x_1, x_2) in the normal and tangent direction to the boundary. The angle θ is defined so that

$\Omega(x,t) = (\cos(\theta(x,t)), \sin(\theta(x,t)))$ in this basis. System (IV.4)-(IV.5) then becomes (the index ε is omitted):

$$\partial_t \rho + \partial_{x_1}(\rho \cos\theta) + \partial_{x_2}(\rho \sin\theta) = 0, \tag{IV.18}$$

$$[\partial_t \theta + (\cos\theta \partial_{x_1}\theta + \sin\theta \partial_{x_2}\theta) + (-\sin\theta \partial_{x_1} \varepsilon p(\rho) + \cos\theta \partial_{x_2} \varepsilon p(\rho))] \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = 0 \tag{IV.19}$$

We suppose that all quantities have locally smooth variations in the direction tangent to the boundary and we focus on the possible sharp variations or discontinuities in the normal direction. To analyze this situation, we perform a coordinate dilation in the x_1 direction and in time: $x_1' = \delta x_1$, $x_2' = \delta x_2$, $t' = \delta t$, with $\delta \ll 1$. In these new variables, all x_1 and t derivatives are multiplied by $1/\delta$. Letting $\delta \to 0$, we are led to the following one-dimensional system with $x_1 = x$:

$$\partial_t \rho + \partial_x(\rho \cos\theta) = 0, \tag{IV.20}$$

$$\partial_t \theta + \cos\theta \partial_x \theta + \varepsilon \sin^2\theta \partial_x \varepsilon p(\rho) = 0. \tag{IV.21}$$

Hyperbolic systems like (IV.20)-(IV.21) have analytical solutions which are those of the Riemann problem. These solutions are associated to initial conditions which consist of a discontinuity between two constant states. We will construct the solutions of the Riemann problem for system (IV.20)-(IV.21) and analyze their limits as $\varepsilon \to 0$. This analysis will give rise to jump conditions at the cluster boundaries for these solutions. We will then postulate that these jump conditions are generic and valid for all solutions of the limit problem (IV.8)-(IV.11).

As underlined above, the non-conservative form of system (IV.20)-(IV.21) induces a lack of information about the jump conditions across a boundary. In order to waive the ambiguity, we have to make further assumptions. One of them is to consider the following conservative system as a way to select discontinuities

$$\partial_t \rho + \partial_x(\rho \cos\theta) = 0, \tag{IV.22}$$

$$\partial_t \Psi(\cos(\theta)) + \partial_x(\Phi(\cos\theta) + \varepsilon p(\rho)) = 0. \tag{IV.23}$$

where $\Psi(\cos\theta) = -\ln|\tan(\theta/2)|$ and $\Phi(\cos\theta) = -\ln|\sin\theta|$. It is one of the simplest conservation forms that system (IV.20)-(IV.21) can take. It is obtained by dividing (IV.21) by $\sin^2\theta$. The functions Ψ and Φ satisfy:

$$\frac{d}{d\theta}(\Psi(\cos\theta)) = \frac{1}{\sin^2\theta}, \quad \frac{d}{d\theta}(\Phi(\cos\theta)) = \frac{\cos\theta}{\sin^2\theta}.$$

Other conservative forms of (IV.20)-(IV.21) do exist (see appendix B) but we consider this form because it is a simple one. Note that this conservative form is not equivalent to the original form (IV.20)-(IV.21) because $\Psi(\cos\theta)$ is an even function of θ. Hence it does not provide information on the sign of θ. However, this information will easily be recovered at the end. We remind that, if all conservative forms are equivalent for smooth solutions, they differ for weak solutions. Therefore, the choice of a particular conservative form must be made on physical considerations. Such physical considerations are not available here. In front of this lack of information, the choice of one of the most simple of these conservative forms seems to be the most natural one.

Classical hyperbolic system theory will enable us to solve the Riemann problem for (IV.22)-(IV.23) and to take the limit $\varepsilon \to 0$ of these solutions. The limit solutions will satisfy some jump relations which we will assume generic of all solutions of the limit problem (IV.8)-(IV.11). We now present the result of this analysis for such generic solutions. We call "unclustered" region (UC) the domains where $0 < \rho < \rho^*$, by contrast to vacuum (V) where $\rho = 0$ or clusters (C) where $\rho = \rho^*$.

Formal Statement 1. *The boundary conditions at cluster boundaries or vacuum boundaries of system (IV.8)-(IV.11) are as follows:*

- **Interface (C)-(UC).** *The pressure jump is given by*

$$[\bar{p}] = \frac{[\Psi(\cos\theta)]\,[\rho\cos\theta]}{[\rho]} - [\Phi(\cos\theta)]\,, \tag{IV.24}$$

  and the shock speed is given by the Rankine-Hugoniot relation

$$\sigma = [\rho\cos(\theta)]/[\rho], \tag{IV.25}$$

  where the angle brackets denote the jumps across the interface. We note that $p_{UC} = 0$ and that specifying $[\bar{p}]$ actually specifies the boundary value of $\bar{p}$ at the cluster boundary.
- **Interface (UC)-(V).** *The interface speed σ is equal to the fluid normal speed $\sigma = \cos\theta = \Omega \cdot n$ at the boundary of the (UC) region*

$$\sigma = (\cos\theta)_{UC}, \tag{IV.26}$$

  and the pressure $\bar{p}$ is identically zero.
- **Interface (C)-(V).** *The interface speed is equal to the normal speed $\cos\theta = \Omega \cdot n$ at the cluster boundary and the boundary value of $\bar{p}$ is zero*

$$\sigma = (\cos\theta)_C, \quad \bar{p}_C = 0. \tag{IV.27}$$

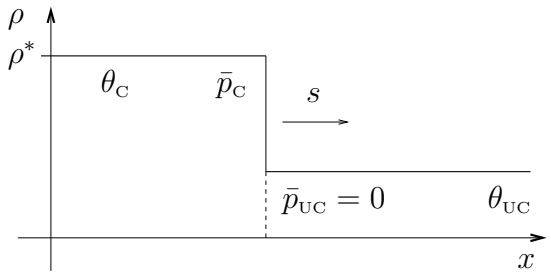
- **Interface (UC)-(UC).** *This is a contact discontinuity between two regions of different ρ. The normal velocity is continuous and equal to the speed of the discontinuity*

$$[\cos\theta] = 0, \quad \sigma = \cos\theta, \tag{IV.28}$$

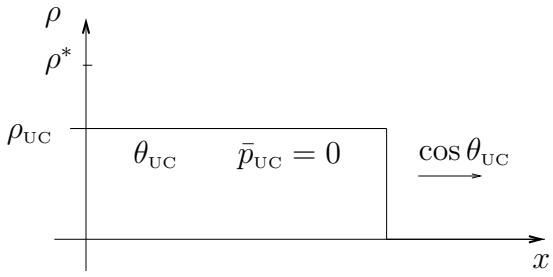
  and the pressure is identically zero.

We note that all these statements are consistent with proposition 13. Section 3 provides the detailed analysis which leads to these relations. The dynamics of the interface between two clusters (C)-(C) does not follow from the analysis of the Riemann problem. We provide a separate analysis of it by introducing the so-called cluster dynamics.

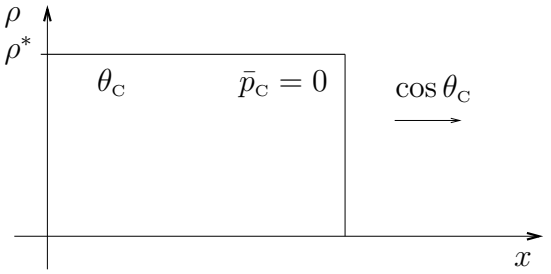
We note that (C)-(UC) interfaces or contact discontinuities (UC)-(UC) may incorporate a flip of the sign of θ in the velocity jump. This has no influence on the boundary values of $\bar{p}$ at the cluster boundary which is the quantity we wish to determine by this analysis. The Formal Statement 1 is illustrated in figure IV.4.



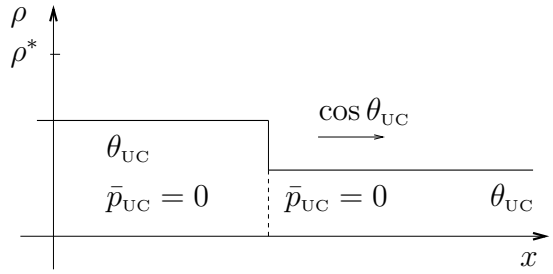
(a) Interface (C)-(UC). $\bar{p}_{\mathrm{C}}$ and s are given by (IV.24)-(IV.25).



(b) Interface (UC)-(V)



(c) Interface (C)-(V)



(d) Interface (UC)-(UC)

Figure IV.4: Interfaces

2.5 Clusters dynamics

We now focus on the interface (C)-(C), i.e. a collision of two clusters. The procedure using limits $\varepsilon \to 0$ of the Riemann problem does not lead to any conclusion since the pressure becomes infinite. Note that this is also the case when dealing with the same limit in the standard Euler problem. Therefore, we have to find another strategy than using the Riemann problem. We turn our attention to the collision between two clusters of finite size and we show that the pressure involves a Dirac delta at the time of the collision. Such an analysis is inspired by the sticky block solutions presented in [10].

Consider two one-dimensional clusters which collide at a time t_c (see fig. IV.5). Before collision, the left (resp. right) cluster at time $t < t_c$ extends between $a_\ell(t)$ and $b_\ell(t)$ (resp. $a_r(t)$ and $b_r(t)$) and moves with speed

$$\cos\theta_\ell = a_\ell'(t) = b_\ell'(t) \quad (\text{resp. } \cos\theta_r = a_r'(t) = b_r'(t)),$$

After the collision, the two clusters agregate and form a new cluster at time $t > t_c$ extending between $a(t)$ and $b(t)$ and moving with speed $\cos\theta = a'(t) = b'(t)$. Therfore, ρ and θ are given for $t < t_c$ by

$$\rho = \rho^* 1_{[a_\ell(t),b_\ell(t)]} + \rho^* 1_{[a_r(t),b_r(t)]}, \quad \theta = \theta_\ell 1_{[a_\ell(t),b_\ell(t)]} + \theta_r 1_{[a_r(t),b_r(t)]},$$

and for $t > t_c$ by

$$\rho = \rho^* 1_{[a(t),b(t)]}, \quad \theta = \theta 1_{[a(t),b(t)]}.$$

where 1_I denotes the indicator function of the interval I (i.e. $1_I(x) = 1$ if $x \in I$ and 0 otherwise). We denote by $m = b_\ell(t_c) = a_r(t_c)$ the collision point. We look for a pressure

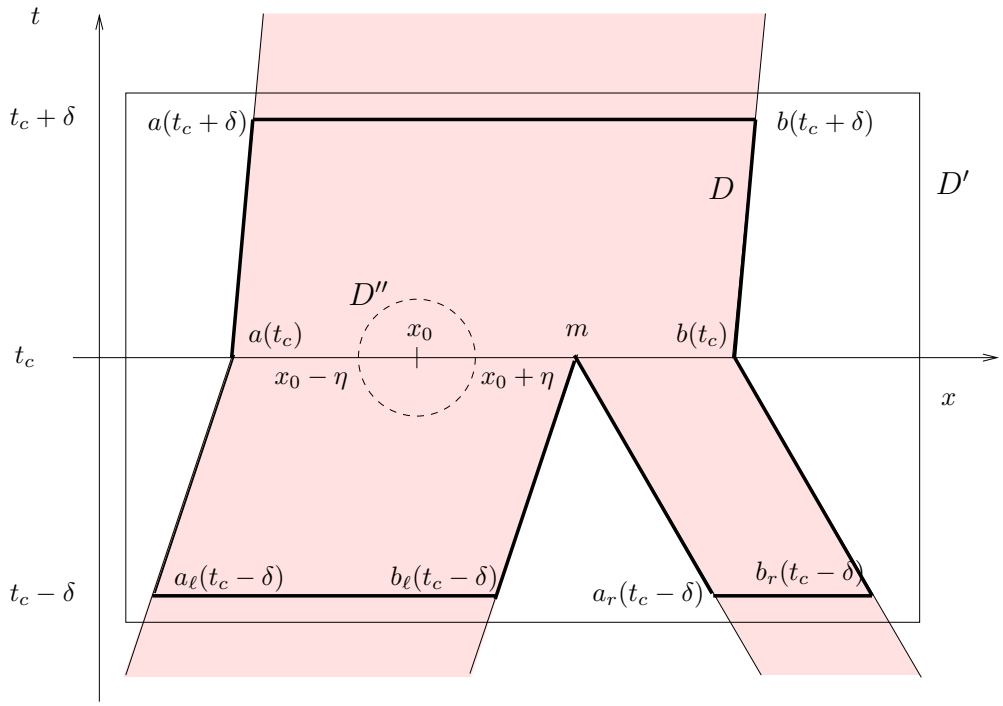


Figure IV.5: Collision of clusters. In the filled domain: clusters ($\rho = \rho^*$).

written as $\bar{p}(x,t) = \pi(x)\delta(t - t_c)$. The following proposition provides conditions for such type of solutions to exist.

Proposition 14. *1- Supposing that $\bar{p}(x,t) = \pi(x)\delta(t - t_c)$ where π is continuous and zero outside the clusters, then θ and π satisfy*

$$(\Psi(\cos\theta) - \Psi(\cos\theta_\ell))(m - a(t_c)) + (\Psi(\cos\theta) - \Psi(\cos\theta_r))(b(t_c) - m) = 0, \tag{IV.29}$$

$$\pi(x) = \begin{cases} (\Psi(\cos\theta) - \Psi(\cos\theta_\ell))(m - x) & \\ \quad +(\Psi(\cos\theta) - \Psi(\cos\theta_r))(b(t_c) - m), & \text{if } x \in [a(t_c), m], \\ (\Psi(\cos\theta) - \Psi(\cos\theta_r))(b(t_c) - x), & \text{if } x \in [m, b(t_c)], \end{cases} \tag{IV.30}$$

2 - Under conditions (IV.29)-(IV.30), (ρ, θ, p) is a solution (in the distributional sense) of (IV.22)-(IV.23).

The proof of this proposition is developed in appendix C.

2.6 Conclusion of the analysis

The underdetermined problem (IV.8)-(IV.11) must be complemented with the Formal Statement 1 which determines the boundary values of $\bar{p}$ at cluster boundaries and by proposition 14 which determines the evolution of two clusters when they meet. Strictly speaking, proposition 14 only gives the collision dynamics of two clusters in dimension 1. In dimension 2, clusters may have complicated shapes. So, the collision dynamics of two clusters in dimension 2 is a complicated problem which will be examined in a future

work. At the present stage, problem (IV.8)-(IV.11) complemented with statement 1 fully determines the dynamics of the limit system as long as two clusters do not meet.

A rigorous theory of the well-posedness of system (IV.8)-(IV.11) complemented with statement 1 is outside the scope of the present paper. Let us just mention how a time discretized version of the problem can be computed. Suppose that $\rho^n(x)$, $\Omega^n(x)$, $\bar{p}^n(x)$ are approximations of $\rho(x,t^n)$, $\Omega(x,t^n)$, $\bar{p}(x,t^n)$ at time $t^n = n\Delta t$. We solve the implicit system

$$\frac{\rho^{n+1}-\rho^n}{\Delta t} + \nabla_x \cdot (\rho^{n+1}\Omega^{n+1}) = 0,$$
$$\frac{\Omega^{n+1}-\Omega^n}{\Delta t} + (\Omega^n \cdot \nabla_x)\Omega^{n+\frac{1}{2}} + (\mathrm{Id} - \Omega^{n+\frac{1}{2}} \otimes \Omega^{n+\frac{1}{2}})\nabla_x \bar{p}^{n+1} = 0,$$

with

$$\Omega^{n+\frac{1}{2}} = \frac{\Omega^n + \Omega^{n+1}}{|\Omega^n + \Omega^{n+1}|}.$$

This form guarantees that $|\Omega^{n+1}|^2 = |\Omega^n|^2 = 1$ (by taking the dot product by $\Omega^{n+\frac{1}{2}}$ and using that $|\Omega^{n+\frac{1}{2}}| = 1$). $\bar{p}^{n+1}$ is determined by solving the elliptic equation

$$-\nabla_x \cdot ((\mathrm{Id} - \Omega^{n+\frac{1}{2}} \otimes \Omega^{n+\frac{1}{2}})\nabla_x \bar{p}^{n+1}) = \nabla_x \cdot ((\Omega^n \cdot \nabla_x)\Omega^{n+\frac{1}{2}}),$$

on every connected component of the cluster region defined at time t by $\{x \in \mathbb{R}^2 \,|\, \rho^{n+1}(x,t) = \rho^*\}$. This equation must be supplemented with suitable boundary conditions on $\bar{p}$ at the boundary of the cluster. These boundary conditions are actually given by the Formal Statement 1, with right-hand sides evaluated at time t^{n+1}. The resolution of this equation guarantees that $\nabla \cdot \Omega^{n+1} = 0$ on every connected component of a cluster, and shows that $\rho^{n+1} = \rho^n = \rho^*$ on such a cluster. Of course, the implicitness of the discretization leads to a nonlinear stationary problem, and the question of the existence of solutions for such a problem is not clear. However, intuitively, it seems that the prescription of the boundary values of $\bar{p}$ at cluster boundaries through the Formal Statement 1 leads to a well-posed problem, at least as long as two clusters do not meet.

3 The one-dimensional Riemann Problem

3.1 Methodology

To find out jump relations satisfied by the solutions of the system (IV.8)-(IV.11), the strategy is to solve the Riemann problem of the one-dimensional perturbation system (IV.22)-(IV.23) and to take the limit $\varepsilon \to 0$ of its solutions. This strategy was successfully adopted for a model of traffic jams in [6].

We note that the eigenvalues and eigenvectors of the hyperbolic system (IV.22)-(IV.23) are

$$\lambda^\varepsilon_\pm(\rho,\theta) = \cos\theta \pm \sqrt{\varepsilon p'(\rho)\rho}|\sin\theta|, \quad r^\varepsilon_\pm(\rho,\theta) = \begin{pmatrix} \pm\rho|\sin\theta| \\ \sqrt{\varepsilon p'(\rho)\rho} \end{pmatrix}.$$

In this conservative system, the domain of θ is restricted to the interval $]0,\pi[$. But this is not a problem since our main concern is to find the missing conditions on $\bar{p}$ at the cluster boundary, and these only depend on jump conditions as functions of $\cos\theta$.

3.2 Solutions to the Riemann problem for (IV.22)-(IV.23)

3.2.1 Genuinely nonlinear fields

The Lax theorem provides the local entropic solutions of the Riemann problem provided that all the fields are totally genuinely nonlinear ($\nabla\lambda^\varepsilon_\pm \cdot r^\varepsilon_\pm \neq 0$) or totally linearly degenerate ($\nabla\lambda^\varepsilon_\pm \cdot r^\varepsilon_\pm = 0$). Unfortunately, the following result implies that the fields are genuinely nonlinear except on a one-dimensional manifold.

Proposition 15. *1. The linearly degenerate set ($\nabla\lambda^\varepsilon_\pm \cdot r^\varepsilon_\pm = 0$) consists of two curves $\mathcal{C}^\varepsilon_\pm$ (each of them corresponds to one characteristic field):*

$$\mathcal{C}^\varepsilon_\pm = \{(\rho,\theta)\,,\ \rho \in [0,\rho^*[,\ \ cotan\theta = \mp G^\varepsilon(\rho)\}\,,$$

where

$$G^\varepsilon(\rho) := \frac{1}{\sqrt{\varepsilon}}\frac{(p''(\rho)\rho + 3p'(\rho))\rho}{(p'(\rho)\rho)^{3/2}} \underset{\rho\to\rho^*}{\sim} C\frac{(\rho^*-\rho)^{\frac{\gamma-1}{2}}}{\sqrt{\varepsilon}}.$$

2. For $\gamma = 1$, the linearly degenerate set tends to the straight lines $\{\theta = 0\}$ and $\{\theta = \pi\}$ as ε tends to 0. For $\gamma > 1$, $\mathcal{C}^\varepsilon_+$ (resp. $\mathcal{C}^\varepsilon_-$) is a one to one, onto mapping from $[0,\rho^*]$ to $[\pi/2,\pi]$ (resp. $[0,\pi/2]$) for all ε, called $\theta^\varepsilon_{ld,\pm}(\rho)$. For a fixed $\theta \in]0,2\pi[$, the inverse map $\rho^\varepsilon_{ld,\pm}(\theta)$ satisfies: $\rho^* - \rho^\varepsilon_{ld,\pm}(\theta) = O(\varepsilon^{\frac{1}{\gamma-1}})$.

The proof of this proposition is easy and is omitted. Thus, the Lax theorem is valid at least locally in the neighbourhood of all the states except those which are on the one-dimensional manifolds. The second part of the previous proposition shows that all the states have locally genuinely nonlinear fields as ε tends to 0. Indeed, even if the state converges to a congested state as $\varepsilon \to 0$, its convergence is like $O(\varepsilon^{1/\gamma})$ (cf. (IV.7)), which is slower than the convergence of the linearly degenerate field when $\gamma > 1$. Therefore, there exists ε' such that for all $\varepsilon < \varepsilon'$ the fields of the converging state are genuinely non-linear. It is also trivially the case when γ equals 1. According to standard nonlinear conservation theory [43], (for ε small enough) the solutions of the Riemann problem consist of two simple waves (shock waves and/or rarefaction waves) of the first and second characteristic fields, separated by constant states.

3.2.2 Shock and rarefaction waves.

A **shock wave** between two constant states (ρ_ℓ,θ_ℓ) and (ρ_r,θ_r) travelling with a constant speed σ satisfies the Rankine-Hugoniot relations:

$$[\rho\cos(\theta)] = \sigma\,[\rho]\,, \tag{IV.31}$$

$$[\Phi(\cos(\theta)) + \varepsilon p(\rho)] = \sigma\,[\Psi(\cos(\theta))]\,, \tag{IV.32}$$

where $[f] := f_r - f_\ell$ denotes the difference between the right value and the left value of any quantity f. By eliminating σ in these equations, we get a non-linear relation between the left and right states:

$$H_\varepsilon(\rho_\ell,\theta_\ell,\rho_r,\theta_r) := [\Phi(\cos(\theta)) + \varepsilon p(\rho)]\,[\rho] - [\Psi(\cos(\theta))]\,[\rho\cos(\theta)] = 0. \tag{IV.33}$$

With a fixed left state, the zero set of H_ε is called the Hugoniot locus and represents all the admissible right states, connected to this left state by a shock wave.

Proposition 16. *The Hugoniot locus consists of two Hugoniot curves $\mathcal{H}^\varepsilon_\pm$ associated to the two caracteristic fields.*

1. *The Hugoniot curve $\mathcal{H}^\varepsilon_-$ associated to λ^ε_- (resp. $\mathcal{H}^\varepsilon_+$ to λ^ε_+) is strictly increasing (resp. strictly decreasing) in the (ρ,θ)-plane. Let $h^\varepsilon_- :](h^\varepsilon_-)^{-1}(0),\pi[\to[0,\rho^*[$ and $h^\varepsilon_+ :]0,(h^\varepsilon_+)^{-1}(0)[\to[0,\rho^*[$ be the Hugoniot curves as functions of θ on their domains of definition.*
2. *The Hugoniot locus tends to the union of the straight lines $\{\theta=\theta_\ell\}$ and $\{\rho=\rho^*\}$.*

The proof of this proposition is developed in appendix D.

A **rarefaction wave** is a continuous self-similar solution $(\rho(\frac{x}{t}),\theta(\frac{x}{t}))$. It satisfies the diffential equation

$$\begin{pmatrix}\rho'(s)\\ \eta'(s)\end{pmatrix}=\frac{r^\varepsilon_\pm(\rho(s),\eta(s))}{\nabla\lambda^\varepsilon_\pm(\rho(s),\eta(s))\cdot r^\varepsilon_\pm(\rho(s),\eta(s))},$$

where $\eta=\Psi(\cos(\theta))$ is the conservative unknown. Therefore, (ρ,η) belong to the integral curve of $r^\varepsilon_\pm$. By changing the parametrization of the integral curve, we obtain

$$\rho'=\pm\rho|\sin\theta|,\quad \theta'=-\sqrt{\varepsilon p'(\rho)\rho}|\sin\theta|,$$

and then the following integral equation

$$\theta-\theta_\ell=\mp\int_{\rho_\ell}^{\rho}\sqrt{\frac{\varepsilon p'(u)}{u}}du. \tag{IV.34}$$

It defines two integral curves $\mathcal{O}^\varepsilon_\pm$ issued from the state (ρ_ℓ,θ_ℓ). The following proposition summarizes their main properties.

Proposition 17. 1. *The integral curve $\mathcal{O}^\varepsilon_-$ of r^ε_- (resp. $\mathcal{O}^\varepsilon_+$ of r^ε_+) is strictly increasing (resp. stricly decreasing) in the $(\rho,\theta)-$plane. Let $i^\varepsilon_- :](i^\varepsilon_-)^{-1}(0),\pi[\to[0,\rho^*[$ and $i^\varepsilon_+ :]0,\ (i^\varepsilon_+)^{-1}\ (0)[\to[0,\rho^*[$ the rarefaction curves as functions of θ on their domains of definition.*

2. *For all $\gamma\geq 1$, the rarefaction curves tend to the union of the straight lines $\{\theta=\theta_\ell\}$ and $\{\rho=\rho^*\}$. Moreover, for $\theta\in]\theta_\ell,\pi[$ (resp. $\theta\in]0,\theta_\ell[$), $\rho^*-i^\varepsilon_-(\theta)=O(\varepsilon^{\frac{1}{\gamma-1}})$ (resp. $\rho^*-i^\varepsilon_+(\theta)=O(\varepsilon^{\frac{1}{\gamma-1}})$).*

3. *Suppose that the state ρ^ε_ℓ is such that $\rho^\varepsilon_\ell\to\rho^*$ and $\varepsilon p(\rho^\varepsilon_\ell)\to\bar{p}_\ell$. For all $\rho<\rho^\varepsilon_\ell$, $(i^\varepsilon_\pm)^{-1}(\rho)$ satisfies:*

$$|(i^\varepsilon_\pm)^{-1}(\rho)-\theta_r|\leq|(i^\varepsilon_\pm)^{-1}(0)-\theta_r|=O(\varepsilon^{\frac{1}{2\gamma}}).$$

The proof is developed in appendix E.

Entropy conditions. In order to satisfy the Lax entropy condition, each Hugoniot curve $\mathcal{H}^\varepsilon_\pm$ is restricted to right states which have a smaller associated eigenvalue than the left state.

Proposition 18. *The eigenvalue λ^ε_- (resp. λ^ε_+) is a decreasing function of ρ on the Hugoniot curve $\mathcal{H}^\varepsilon_-$ (resp. an increasing function of ρ on $\mathcal{H}^\varepsilon_+$) for $\theta < cotan^{-1}\left((-1/\sqrt{\varepsilon p'(\rho)\rho}\right)$ (resp. $\theta > cotan^{-1}\left(1/\sqrt{\varepsilon p'(\rho)\rho}\right)$).*

Proof. Let $g : \rho \in [0, \rho^*] \to g(\rho) \in [0, \pi]$ be an arbitrary function. The variation of $\lambda^\varepsilon_\pm$ on the graph of g is given by

$$\nabla\lambda^\varepsilon_\pm \cdot \begin{pmatrix} 1 \\ g'(\rho) \end{pmatrix} = \pm(\chi^\varepsilon)'(\rho)\sin\theta + g'(\rho)(-\sin\theta \pm \chi^\varepsilon(\rho)\cos\theta),$$

where $\chi^\varepsilon(\rho) = \sqrt{\varepsilon p'(\rho)\rho}$. Since $\chi'(\rho)$ is positive and the Hugoniot curve $(h^\varepsilon_-)^{-1}$ is increasing, λ^ε_- is a decreasing function of ρ on this curve for $\theta \in]0, \pi[$ such that $(\sin\theta + \chi^\varepsilon(\rho)\cos\theta) > 0$. Similarly, since the Hugoniot curve $(h^\varepsilon_+)^{-1}$ is decreasing, λ^ε_+ is a increasing function of ρ on this curve for $\theta \in]0, \pi[$ such that $(-\sin\theta + \chi^\varepsilon(\rho)\cos\theta) > 0$. ∎

So, in the limit $\varepsilon \to 0$, the reachable right states are those belonging to the upper half-domain. We denote by $S^\varepsilon_\pm = \mathcal{H}^\varepsilon_\pm \cap \left\{(\rho,\theta),\ \lambda^\varepsilon_\pm(\rho,\theta) \le \lambda^\varepsilon_\pm(\rho_\ell,\theta_\ell)\right\}$ the shock curves.

Concerning the integral curves $\mathcal{O}^\varepsilon_\pm$, the admissibility conditions select the curves with increasing eigenvalues and so the curves on the lower half-space. Therefore, the rarefaction curves $R^\varepsilon_\pm$ satisfy $R^\varepsilon_\pm \subset \mathcal{O}^\varepsilon_\pm \cap \{(\rho,\theta), \theta \in]0, \theta_\ell[\}$. The union of the shock and the rarefaction curves form the forward wave curve $W^{f,\varepsilon}_\pm = S^\varepsilon_\pm \cup R^\varepsilon_\pm$, while the union of their complementary sets form the backward wave curve $W^{b,\varepsilon}_\pm = \mathcal{H}^\varepsilon_\pm \backslash S^\varepsilon_\pm \cup \mathcal{O}^\varepsilon_\pm \backslash R^\varepsilon_\pm$.

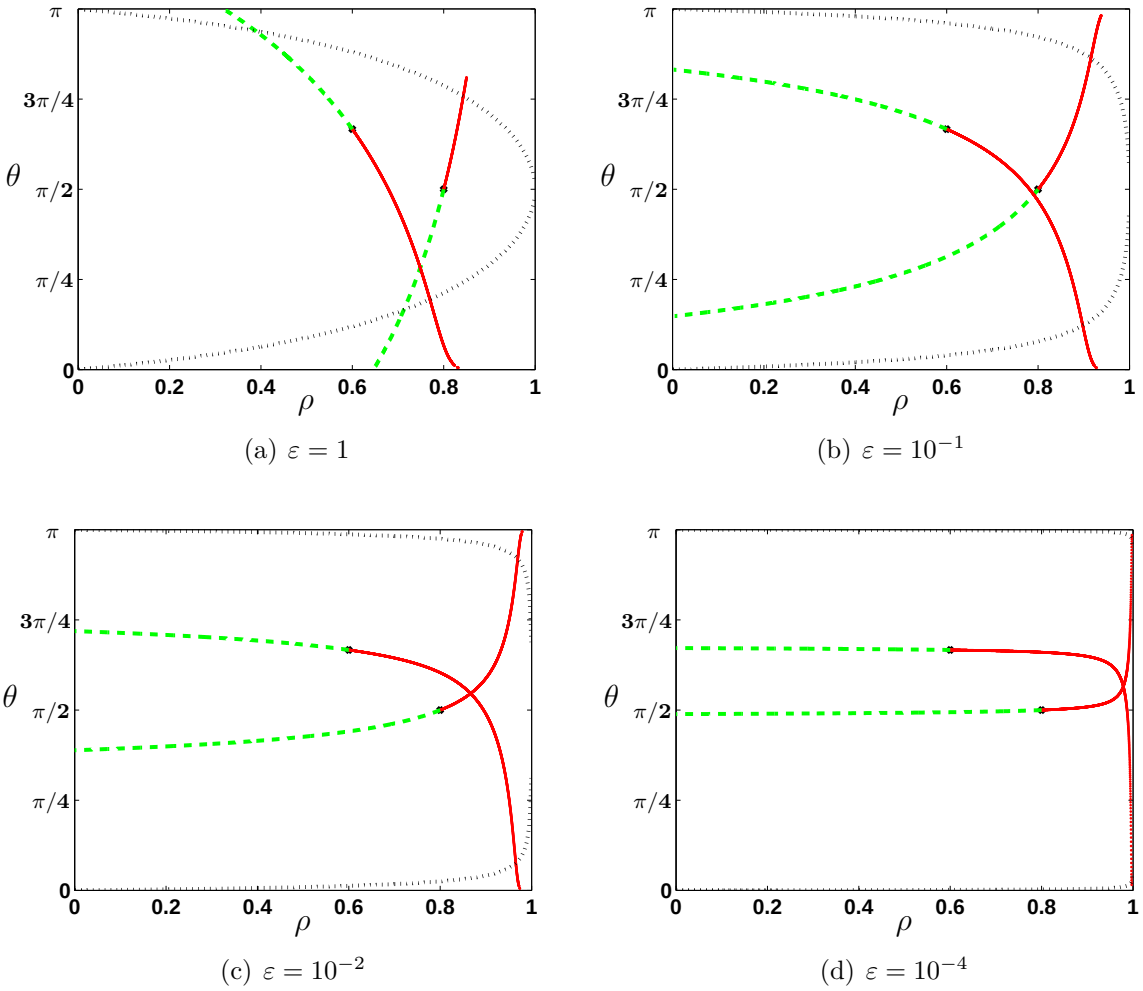
3.2.3 Solutions to the Riemann problem

Given a left state (ρ_ℓ, θ_ℓ) and a right state (ρ_r, θ_r), an entropic solution is found by intersecting the forward 1-wave curve $W^{f,\varepsilon}_-$ issued from the left state and the backward 2-wave curve $W^{b,\varepsilon}_+$ issued from the right state (cf. fig. IV.6). In the following study, the curves indexed by - (resp. by +) are implicitly those issued from the left state (resp. from the right state). Because of the monotony of the shock and rarefaction curves, we can classify the different solutions according to the positions of the left and right states in the (ρ, θ)-plane. The following theorem describes the solution of the Riemann problem for small $\varepsilon > 0$ and is illustrated in figure IV.7.

Theorem 19. *Considering a left state (ρ_ℓ, θ_ℓ) and a right state (ρ_r, θ_r), and for ε small enough, the solution is given by one of the four following cases:*

1. ***Case $\theta_\ell = \theta_r$.*** *If $\rho_\ell < \rho_r$ [resp. $\rho_\ell > \rho_r$], the solution consists of a 1-shock [resp. 1-rarefaction] connecting (ρ_ℓ, θ_ℓ) to $(\widetilde{\rho}, \widetilde{\theta})$ (with $\widetilde{\rho} \in]\rho_\ell, \rho_r[$ and $\widetilde{\theta} > \theta_\ell = \theta_r$ [resp. $\widetilde{\rho} \in]\rho_r, \rho_\ell[$ and $\widetilde{\theta} < \theta_\ell = \theta_r$]) and then a 2-rarefaction [resp. 2-shock] connecting $(\widetilde{\rho}, \widetilde{\theta})$ to (ρ_ℓ, θ_r). This is summarized in the following diagram:*

$$(\rho_\ell,\theta_\ell) \xrightarrow{shock} (\widetilde{\rho},\widetilde{\theta}) \xrightarrow{rarefaction} (\rho_r,\theta_r), \qquad \text{if } \rho_\ell < \rho_r,$$
$$(\rho_\ell,\theta_\ell) \xrightarrow{rarefaction} (\widetilde{\rho},\widetilde{\theta}) \xrightarrow{shock} (\rho_r,\theta_r), \qquad \text{if } \rho_\ell > \rho_r.$$



(a) $\varepsilon = 1$ (b) $\varepsilon = 10^{-1}$

(c) $\varepsilon = 10^{-2}$ (d) $\varepsilon = 10^{-4}$

Figure IV.6: Wave curves $W_{-}^{f,\varepsilon}$ for the left state $(\rho_\ell, \theta_\ell) = (0.8, \pi/2)$ and $W_{+}^{b,\varepsilon}$ for the right state $(\rho_r, \theta_r) = (0.6, 2\pi/3)$. In dashed green lines: the rarefaction curves. In continuous red lines: the shock curves. In dotted black lines: linearly degenerate sets. $\rho^* = 1$, $\gamma = 2$.

2. ***Case*** $\theta_\ell > \theta_r$ ***(***$\cos\theta_\ell < \cos\theta_r$***).*** *The solution consists of a 1-rarefaction connecting* (ρ_ℓ, θ_ℓ) *to* $(0, \widetilde{\theta})$ *(with* $\widetilde{\rho} < \rho_\ell, \rho_r$ *and* $\widetilde{\theta} \in]\theta_r, \theta_\ell[$*) and then a 2-rarefaction wave connecting* $(0, \widetilde{\theta})$ *to* (ρ_r, θ_r)*. We get the following diagram:*

$$(\rho_\ell, \theta_\ell) \overset{rarefaction}{\longrightarrow} (0, \widetilde{\theta}) \overset{vacuum}{\longrightarrow} (0, \widetilde{\widetilde{\theta}}) \overset{rarefaction}{\longrightarrow} (\rho_r, \theta_r).$$

3. ***Case*** $\theta_\ell < \theta_r$ ***(***$\cos\theta_\ell > \cos\theta_r$***).*** *There are two sub-cases:*
   - *if* $\rho_r^\varepsilon < (h_-^\varepsilon)^{-1}(\theta_r^\varepsilon)$ *and* $\rho_\ell^\varepsilon < (h_+^\varepsilon)^{-1}(\theta_\ell^\varepsilon)$*, the solution consists of a 1-shock connecting* (ρ_ℓ, θ_ℓ) *to* $(\widetilde{\rho}, \widetilde{\theta})$ *(with* $\widetilde{\rho} > \rho_\ell, \rho_r$ *and* $\widetilde{\theta} \in]\theta_r, \theta_\ell[$*) and then a 2-shock connecting* $(\widetilde{\rho}, \widetilde{\theta})$ *to* (ρ_r, θ_r)*. The diagram is:*

$$(\rho_\ell, \theta_\ell) \overset{shock}{\longrightarrow} (\widetilde{\rho}, \widetilde{\theta}) \overset{shock}{\longrightarrow} (\rho_r, \theta_r),$$

   - *if* $\rho_r^\varepsilon > (h_-^\varepsilon)^{-1}(\theta_r^\varepsilon)$ *[resp.* $\rho_\ell^\varepsilon > (h_+^\varepsilon)^{-1}(\theta_\ell^\varepsilon)$*], the solution consists of a 1-shock [resp. 1-rarefaction] connecting* (ρ_ℓ, θ_ℓ) *to* $(\widetilde{\rho}, \widetilde{\theta})$ *(with* $\widetilde{\rho} \in]\rho_\ell, \rho_r[$ *and* $\widetilde{\theta} > \theta_r$ *[resp.* $\widetilde{\rho} \in]\rho_r, \rho_\ell[$ *and* $\widetilde{\theta} < \theta_\ell$*]) and then a 2-rarefaction [resp. 2-shock] connecting* $(\widetilde{\rho}, \widetilde{\theta})$ *to* (ρ_r, θ_r)*. The diagram is as follows:*

$$(\rho_\ell, \theta_\ell) \overset{shock}{\longrightarrow} (\widetilde{\rho}, \widetilde{\theta}) \overset{rarefaction}{\longrightarrow} (\rho_r, \theta_r), \qquad \text{if } \rho_\ell < \rho_r,$$

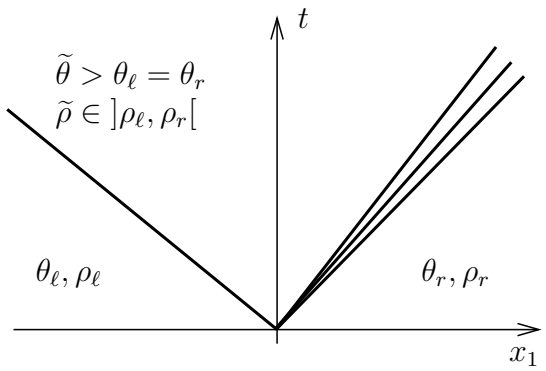
$$(\rho_\ell, \theta_\ell) \overset{rarefaction}{\longrightarrow} (\widetilde{\rho}, \widetilde{\theta}) \overset{shock}{\longrightarrow} (\rho_r, \theta_r), \qquad \text{if } \rho_\ell > \rho_r.$$

The detailed proof of this theorem is developed in appendix F.1. Let us provide some ideas of the proof. For finite ε, there exist four kinds of solutions depending on what parts of the curves $W_-^{f,\varepsilon}$ and $W_+^{b,\varepsilon}$ meet. So, for a fixed left state, the state-space is divided in four subdomains. These subdomains depends on the left state. However, reminding that the limit of the Hugoniot and integral curves are straight lines $\theta = \theta_\ell$ or $\rho = \rho^*$ (cf. propositions 16 and 17) for all left states, the four subdomains have the same behaviour as $\varepsilon \to 0$ whatever the left state is.

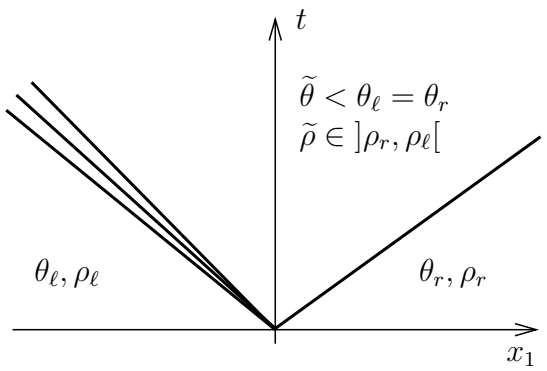
3.2.4 The sign of θ

The conservative system (IV.22)-(IV.23) does not determine the sign of θ (if θ is supposed to be in $]-\pi, \pi[$). As mentioned above, this is not important since our main goal is to provide connection conditions on $\bar{p}$ between the left and right states. However, it is desirable to determine it in the present analysis, for the sake of completeness. For this goal, we cannot use (IV.23) because $\Psi(\cos\theta)$ is an even function of θ. Again, we are facing an indetermination due to the non-conservative character of the system. One possible solution is to introduce a contact discontinuity from θ to $-\theta$ with propagation speed $\cos\theta$ in the domains where ρ is constant and $\cos\theta$ is continuous. If we add such a contact wave, there is only one possible construction given by the following:

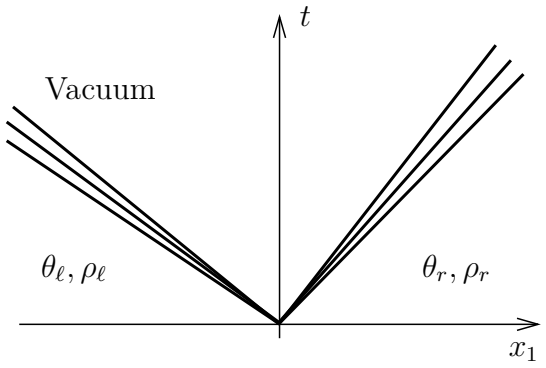
Proposition 20. *Suppose that* θ_ℓ, $\theta_r \in [-\pi, \pi]$ *and* θ_ℓ, θ_r *have different signs.*



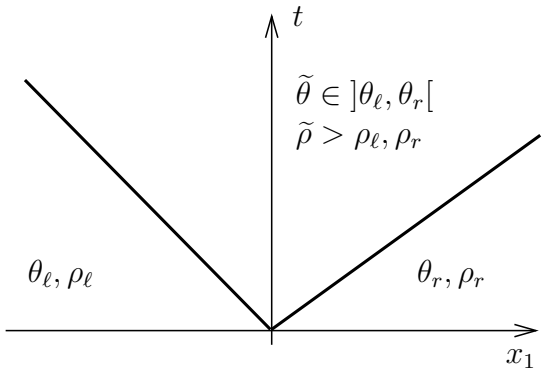
(a) Case $\theta_\ell = \theta_r$, $\rho_\ell < \rho_r$



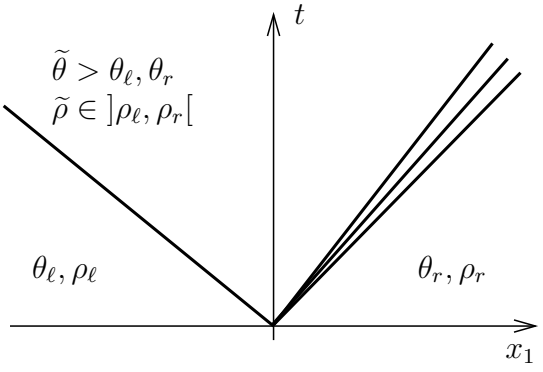
(b) Case $\theta_\ell = \theta_r$, $\rho_\ell > \rho_r$



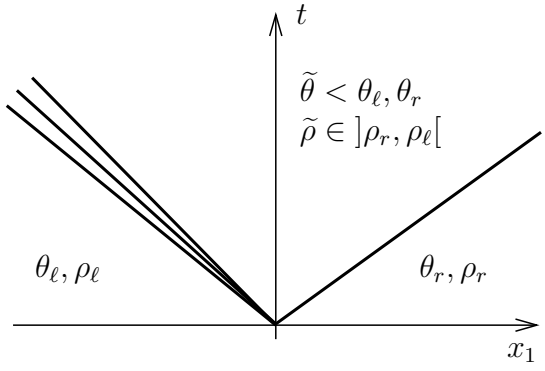
(c) Case $\theta_\ell > \theta_r$



(d) Case $\theta_\ell < \theta_r$, first subcase

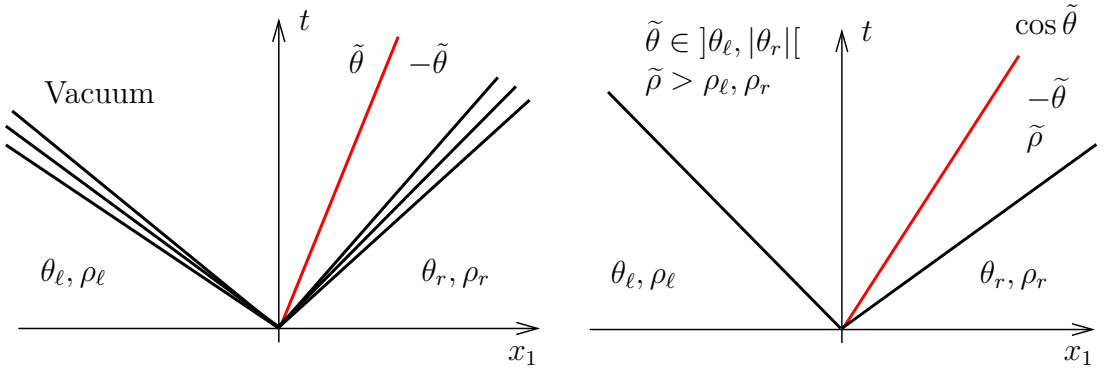


(e) Case $\theta_\ell < \theta_r$, second subcase, $\rho_\ell < \rho_r$



(f) Case $\theta_\ell < \theta_r$, second subcase, $\rho_\ell > \rho_r$

Figure IV.7: Solutions to the Riemann problem for small $\varepsilon > 0$.



(a) Case $\theta_\ell > |\theta_r|$ (b) Case $\theta_\ell < |\theta_r|$, first subcase

Figure IV.8: Some solutions to the Riemann problem for small $\varepsilon > 0$ and $-\pi < \theta_r < 0 < \theta_\ell < \pi$.

1. *In the subcases* $\cos\theta_\ell = \cos\theta_r$ *and* $\cos\theta_\ell > \cos\theta_r$ *of theorem 19, the only possibility for a contact wave connecting two states with angles respectively equal to* $\tilde\theta$ *and* $-\tilde\theta$ *is to propagate in the region of constant* ρ *and continuous* $\cos\theta$ *corresponding to the intermediate state of the conservative system. The propogation speed then equals to* $\cos\tilde\theta$.
2. *In the subcase* $\cos\theta_\ell < \cos\theta_r$, *the only possibility for a contact wave connecting two states with angles respectively equal to* $\tilde\theta$ *and* $-\tilde\theta$ *is to propagate in the region corresponding to vacuum. There is no uniqueness of the propagation speed but since this contact discontinuity occurs within a vacuum region* $\rho = 0$, *we may consider that* θ *is not defined in this region.*

The proof of this proposition can be found in appendix F.2. Two cases of the Riemann problem with $\theta_r < 0 < \theta_\ell$ are represented in Fig. IV.8. Note that the position of the contact wave does not depend on ε. So their limits as ε goes to zero are easily obtained.

3.3 The solutions of the Riemann problem in the limit $\varepsilon \to 0$

In order to study the limit $\varepsilon \to 0$, we introduce converging sequences of left and right states

$$((\rho_\ell^\varepsilon, \theta_\ell^\varepsilon), (\rho_r^\varepsilon, \theta_r^\varepsilon)) \quad \underset{\varepsilon\to 0}{\longrightarrow} \quad ((\rho_\ell, \theta_\ell), (\rho_r, \theta_r)),$$

and we look for the limits of the solutions of the associated Riemann problems. There are three cases to consider: either none of the two states tends to the congested state $(\rho_\ell, \rho_r < \rho^*)$, or one of the two does $(\rho_\ell < \rho^*, \rho_r^\varepsilon \to \rho^*)$ or both of them do $(\rho_\ell^\varepsilon, \rho_r^\varepsilon \to \rho^*)$. The case $(\rho_\ell^\varepsilon \to \rho^*, \rho_r < \rho^*)$ is obtained by symmetry from the case $(\rho_\ell < \rho^*, \rho_r^\varepsilon \to \rho^*)$: the left and right quantities have to be exchanged and the arrows have to be flipped (like in the first case of theorem 19). Since the solutions of the Riemann problem are bounded and monotonous, all the sequences belong to a bounded subset of $BV(\mathbb{R})$ and consequently, to a compact subset of $L^1_{\text{loc}}(\mathbb{R})$. So we only need to prove the uniqueness of the limit of converging sequences to prove the convergence of the whole sequence and we

can consider that the convergence is in the almost everywhere sense (up to the extraction of a subsequence).

As a guideline, we mention that, compared with the system with finite ε, the limit Riemann problem has two additional properties: the appearance of clusters which corresponds to the saturation of the constraint $\rho \leq \rho^*$ and the disappearance of rarefaction waves and their transformations into contact waves. In the subsequent statements, the term "limit" is a short-hand for "limit of the solution to the Riemann problem of (IV.22)-(IV.23)" as $\varepsilon \to 0$.

3.3.1 Case $\rho_\ell < \rho^*, \rho_r < \rho^*$ (see fig. IV.9)

Proposition 21. ***(Case*** $\rho_\ell < \rho^*, \rho_r < \rho^*$***)*** *There are only three cases:*

(a) ***Subcase*** $\theta_\ell = \theta_r$***.*** *The limit consists of only one contact wave connecting* (ρ_ℓ, θ_ℓ) *to* (ρ_r, θ_ℓ)*:*

$$(\rho_\ell, \theta_\ell) \quad \overset{contact}{\longrightarrow} \quad (\rho_r, \theta_r).$$

The travelling speed is equal to $\cos\theta_\ell$*.*

(b) ***Subcase*** $\theta_\ell > \theta_r$***.*** *The limit consists of two contact waves connecting the two states to a vacuum state:*

$$(\rho_\ell, \theta_\ell) \quad \overset{contact}{\longrightarrow} \quad (0, \theta_\ell) \quad \overset{Vacuum}{\longrightarrow} \quad (0, \theta_r) \quad \overset{contact}{\longrightarrow} \quad (\rho_r, \theta_r).$$

The travelling speeds are respectively equal to $\cos\theta_\ell$ *and* $\cos\theta_r$*.*

(c) ***Subcase*** $\theta_\ell < \theta_r$***.*** *The limit consists of two shocks connecting the left state* (ρ_ℓ, θ_ℓ) *to a congested state* $(\rho^*, \widetilde{\theta}, \bar{p})$ *and then connecting* $(\rho^*, \widetilde{\theta}, \bar{p})$ *to the right state* (ρ_r, θ_r)*:*

$$(\rho_\ell, \theta_\ell) \quad \overset{shock}{\longrightarrow} \quad (\rho^*, \widetilde{\theta}, \bar{p}) \quad \overset{shock}{\longrightarrow} \quad (\rho_r, \theta_r).$$

where $\widetilde{\theta}$ *is the unique solution of*

$$\begin{aligned} &[\Psi(\cos(\theta))]_r \frac{[\rho\cos(\theta)]_r}{[\rho]_r} - [\Psi(\cos(\theta))]_\ell \frac{[\rho\cos(\theta)]_\ell}{[\rho]_\ell} = [\Phi(\cos(\theta))]_r^\ell, \\ &\theta \in [\min(\theta_\ell, \theta_r), \max(\theta_\ell, \theta_r)], \end{aligned}$$

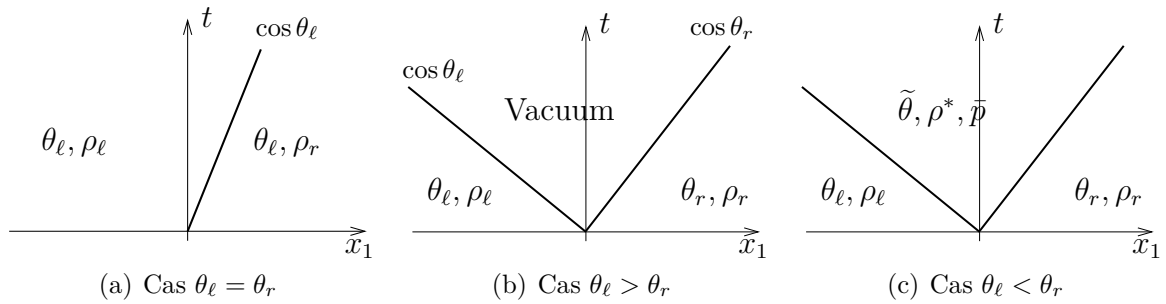
and $\bar{p}$ *is given by*

$$\bar{p} = \frac{[\Psi(\cos(\theta))]_\ell [\rho\cos(\theta)]_\ell}{[\rho]_\ell} - [\Phi(\cos(\theta))]_\ell = \frac{[\Psi(\cos(\theta))]_r [\rho\cos(\theta)]_r}{[\rho]_r} - [\Phi(\cos(\theta))]_r.$$

The shock speeds are given by the Rankine-Hugoniot condition for the density (IV.31).

Note that in clustered region, since $\rho = \rho^*$, the state is determined by the values of θ and $\bar{p}$. This is why we add a third component giving the value of $\bar{p}$ to the vector defining the state in the clustered region.

In this proposition, the quantities $[f]_\ell := \widetilde{f} - f_\ell$, $[f]_r := \widetilde{f} - f_r$ denote the difference between the intermediate value and the left (or right) value of the quantity f and $[f]_\ell^r := f_r - f_\ell$ denotes the difference between the right and left values of the quantity f.



(a) Cas $\theta_\ell = \theta_r$ (b) Cas $\theta_\ell > \theta_r$ (c) Cas $\theta_\ell < \theta_r$

Figure IV.9: Limit solutions of the Riemann problem for $\varepsilon = 0$ and $\rho_\ell, \rho_r < \rho^*$.

This proposition covers several kinds of interfaces described in Formal Statement 1: the case (a) is an occurence of an interface (UC)-(UC), the case (b) of an interface (UC)-(V) and the case (c) of an interface (C)-(UC). Moreover, the proposition implies that θ is continuous inside (UC) domains.

Note that all the intermediate states are explicitly given or are solutions of a non-linear equation and so, are explicitly computable. The proof of this proposition is given in appendix G.1.

3.3.2 Case $\rho_\ell < \rho^*, \rho_r = \rho^*$ (see fig. IV.10)

This case is typical of the situation at a cluster boundary. In this case, the main new feature is the appearance of declustering waves as limits of the rarefaction waves. These declustering waves are instantaneous cancellations of the pressure. The following lemma details this statement:

Lemma 22. *(Limit of rarefaction waves, declustering wave) Let $(\rho_r^\varepsilon, \theta_r)$ be a sequence of right states such that $\rho_r^\varepsilon \to \rho^*$ and $\varepsilon p(\rho_r^\varepsilon) \to \bar{p}_r > 0$. Introduce a converging sequence of states $(\widetilde{\rho}^\varepsilon, \widetilde{\theta}^\varepsilon)$ lying on the rarefaction curves issued from the right states, such that $\widetilde{\rho}^\varepsilon < \rho_r^\varepsilon$.*

If $\widetilde{\rho} = \lim \widetilde{\rho}^\varepsilon < \rho^*$, then the rarefaction wave tends to the combination of a contact wave between the state $(\widetilde{\rho}, \theta_r)$ and $(\rho^*, \theta_r, \bar{p}_r)$ with speed $\cos\theta_r = \lambda_+$ and a declustering wave, i.e. a contact wave with infinite speed which cancels the pressure, which provides a transition between $(\rho^*, \theta_r, \bar{p}_r)$ and $(\rho^*, \theta_r, 0)$.

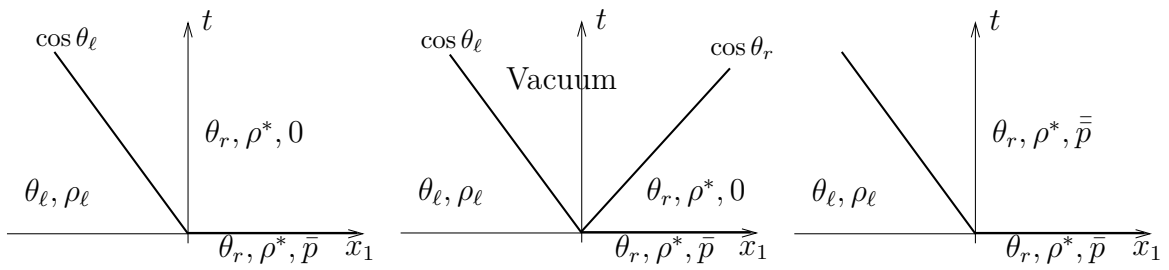
If $\widetilde{\rho} = \lim \widetilde{\rho}^\varepsilon = \rho^*$ then the rarefaction wave tends to a shock wave with infinite speed between the states $(\rho^*, \theta_r, \bar{\bar{p}})$ and $(\rho^*, \theta_r, \bar{p}_r)$, where $\bar{\bar{p}} = \lim \varepsilon p(\widetilde{\rho}^\varepsilon)$.

The proof of this lemma is developed in appendix G.2.

The next proposition provides the solutions of the limit Riemann problem and figure IV.10 schematically describes them.

Proposition 23. ***(Case** $\rho_\ell < \rho^*, \rho_r = \rho^*$**)** *There are only three cases:*

(a) ***Subcase** $\theta_\ell = \theta_r$**.** *The limit solution consists of one contact wave connecting the left state (ρ_ℓ, θ_ℓ) to an intermediate congested state $(\rho^*, \theta_r, \bar{p} = 0)$ and then a cluster*



(a) Case $\theta_\ell = \theta_r$ (b) Case $\theta_\ell > \theta_r$ (c) Case $\theta_\ell < \theta_r$

Figure IV.10: Limit solutions of the Riemann problem for $\varepsilon = 0$ and $\rho_\ell < \rho^*$, $\rho_r^\varepsilon \to \rho^*$.

contact (with infinite speed):

$$(\rho_\ell, \theta_\ell) \overset{contact}{\longrightarrow} (\rho^*, \theta_r, 0) \overset{declust.}{\to} (\rho^*, \theta_r, \bar{p}).$$

(b) ***Subcase*** $\theta_\ell > \theta_r$***.*** *The limit solution consists of one contact wave connecting the left state to vacuum, and then another contact wave connecting the vacuum to a congested and pressureless state* $(\rho^*, \theta_r, 0)$ *and finally a cluster contact connecting* $(\rho^*, \theta_r, 0)$ *to* $(\rho^*, \theta_r, \bar{p})$*:*

$$(\rho_\ell, \theta_\ell) \overset{contact}{\longrightarrow} (0, \theta_\ell) \overset{vacuum}{\longrightarrow} (0, \theta_r) \overset{contact}{\longrightarrow} (\rho^*, \theta_r, 0) \overset{declust.}{\to} (\rho^*, \theta_r, \bar{p}).$$

(c) ***Subcase*** $\theta_\ell < \theta_r$***.*** *The limit solution consists of one shock wave connecting the left state* (ρ_ℓ, θ_ℓ) *to an intermediate congested state* $(\rho^*, \theta_r, \bar{\bar{p}})$ *and one contact wave with infinite propagation speed connecting* $(\rho^*, \theta_r, \bar{\bar{p}})$ *to the right state* $(\rho^*, \theta_r, \bar{p})$*:*

$$(\rho_\ell, \theta_\ell) \overset{shock}{\longrightarrow} (\rho^*, \theta_\ell, \bar{\bar{p}}) \overset{contact}{\longrightarrow} (\rho^*, \theta_r, \bar{p}),$$

where the intermediate pressure $\bar{\bar{p}}$ *is equal to*

$$\bar{\bar{p}} = [\Psi(\cos\theta)]_\ell \frac{[\rho\cos\theta]_\ell}{[\rho]_\ell} - [\Phi(\cos\theta)]_\ell .$$

The shock speed is given by the Rankine-Hugoniot condition for the density (IV.31).

In practice, when instantaneous waves occur (i.e. with infinite propagation speed), it means that the initial data of the Riemann problem does not spontaneously appear during the dynamical evolution of the limit problem. They have to be ignored.

Like proposition 21, this new proposition covers several kinds of interfaces described in the Formal Statement 1: cases (a) and (c) are occurences of interfaces (C)-(UC), the left wave of case (b) is an occurence of an interface (UC)-(V) whereas the right wave of the case (b) is an interface (C)-(V).

The proof of this proposition is in appendix G.3.

3.3.3 Case $\rho_\ell = \rho_r = \rho^*, \rho_\ell^\varepsilon < \rho_r^\varepsilon$ (see fig. IV.11)

We assume in addition that $\varepsilon p(\rho_\ell^\varepsilon)$ and $\varepsilon p(\rho_r^\varepsilon)$ have finite positive limits, denoted by $\bar{p}_\ell > 0$ and $\bar{p}_r > 0$. Figure IV.11 provides a sketch of the solutions.

Proposition 24. ***(Case** $\rho_\ell = \rho_r = \rho^*, \rho_\ell^\varepsilon < \rho_r^\varepsilon$**)** *There are only three cases:*

(a) ***Subcase** $\theta_\ell = \theta_r$**.** *The limit solution consists of a uniform constant state $(\rho^*, \theta_\ell, \bar{p}_\ell)$.*

(b) ***Subcase** $\theta_\ell > \theta_r$**.** *The limit solution consists of two contact waves and two cluster contact with infinite travelling speed:*

$$(\rho^*, \theta_\ell, \bar{p}_\ell) \overset{declust.}{\longrightarrow} (\rho^*, \theta_\ell, 0) \overset{contact}{\longrightarrow} (0, \theta_\ell) \overset{vacuum}{\longrightarrow} (0, \theta_r) \overset{contact}{\longrightarrow} (\rho^*, \theta_r, 0) \overset{declust.}{\longrightarrow} (\rho^*, \theta_r, \bar{p}_r),$$

(c) ***Subcase** $\theta_\ell < \theta_r$**.** *The limit solution consists of two shock waves with infinite propagation speed connecting the left state $(\rho^*, \theta_\ell, \bar{p}_\ell)$ to $(\rho^*, \widetilde{\theta}, +\infty)$ and then $(\rho^*, \widetilde{\theta}, +\infty)$ to $(\rho^*, \theta_r, \bar{p}_r)$:*

$$(\rho^*, \theta_\ell, \bar{p}_\ell) \overset{shock}{\longrightarrow} (\rho^*, \widetilde{\theta}, +\infty) \overset{shock}{\longrightarrow} (\rho^*, \theta_r, \bar{p}_r),$$

where $\widetilde{\theta}$ is the only solution of

$$\frac{[\Psi(\cos(\theta))]_r[\cos(\theta)]_r}{[\Psi(\cos(\theta))]_\ell[\cos(\theta)]_\ell} = \left(\frac{\bar{p}_\ell}{\bar{p}_r}\right)^{\frac{1}{\gamma}}. \tag{IV.35}$$

These solutions display only one kind of interface among those discussed in the Formal Statement 1: the case (b) is an occurence of an interface (C)-(V). According to the cases (a) and (c), the solution inside clusters is continuous. However the case (c) does not provide a meaningful solution since the pressure becomes infinite and this is why Formal Statement 1 does not allow to decide what happens at the interface (C)-(C), i.e. a collision of two clusters. It seems to result from the fact that in this case the Riemann problem models the collision of two infinite one-dimensional clusters. Section 2.5 provides a description of the collision between finite-size one-dimensional clusters. We have seen that the pressure $\bar{p}$ involves a Dirac delta in time. Indeed, according to case (c), the infinite propagation speed of the waves inside clusters implies the discontinuity of the function θ in time: $\theta = \theta_\ell + (\widetilde{\theta} - \theta_\ell)H(t - t_c)$, where H denotes here the Heaviside function. Then, equation (IV.23) leads to

$$(\Psi(\widetilde{\theta}) - \Psi(\theta_\ell))\delta(t - t_c) = \partial_x \bar{p},$$

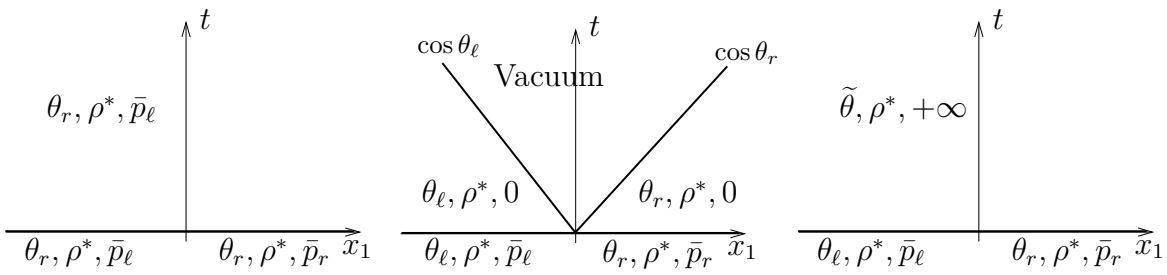
which justifies to look for a pressure with a dirac delta in time. The Riemann problem does not allow to take into account such a pressure.

The proof of the proposition is deferred appendix G.4.

3.4 Connecting the Riemann problem analysis to the Formal Statement 1

We remind that, by contrast with the finite ε system (IV.4)-(IV.6), which is a standard hyperbolic system, the limit system (IV.8)-(IV.11) exhibits two additional characteristics:

(i) the appearance of clusters corresponding to the saturation of the constraint $\rho = \rho^*$,



(a) Case $\theta_\ell = \theta_r$ (b) Case $\theta_\ell > \theta_r$ (c) Case $\theta_\ell < \theta_r$

Figure IV.11: Limit solutions of the Riemann problem for $\varepsilon = 0$ and $\rho_\ell^\varepsilon, \rho_r^\varepsilon \to \rho^*$, $\rho_\ell^\varepsilon < \rho_r^\varepsilon$.

(ii) the appearance of vacuum.

The previous study helps in understanding the dynamics of the interfaces between unclustered states (UC) $0 < \rho < \rho^*$, vacuum (V) $\rho = 0$ and clusters (C) $\rho = \rho^*$. Up to now, no rigorous theory for the limit $\varepsilon \to 0$ exists and so, we cannot have access to these dynamics rigorously. Our method is to investigate these dynamics through the inspection of the limit $\varepsilon \to 0$ of the solutions of the Riemann problem of the finite ε system.

The first remark is that waves with infinite speed correspond to an instantaneous transition from the initial data to some different solution. To some extent, this means that the corresponding initial datum is unstable, and therefore, that it will never appear spontaneously in the course of the evolution of the system. Therefore, we can discard initial data which exhibit this phenomenon, and replace them by the one which is found after the infinite speed wave has been applied.

The second remark is that the various solutions of the Riemann problem can be grouped by situations corresponding to the four cases listed in the formal statement 1, i.e. interfaces (C)-(UC), (UC)-(V), (C)-(V) and (UC)-(UC). As discussed in the lines following statement 24, the case (C)-(C) is not accessible by the Riemann problem analysis because the collision dynamics of two clusters depend on their size, and because the Riemann problem only allows to consider infinite size clusters. This is why cluster collisions are analyzed separately in proposition 14. However, cluster collisions is a complex phenomenon in 2D and proposition 14 only provides a one-dimensional analysis. The two-dimensional analysis is still in progress.

To highlight the link between the Riemann problem analysis and the Formal Statement 1, we point out which solutions of the Riemann problem correspond to which case in the Formal Statement 1:

Interface (C)-(UC): it appears in

- prop. 21, subcase (c): we see that the intermediate congested state is separated from the left and right states by two (C)-(UC) interfaces. We notice that (IV.24) and (IV.25) are respectively the relations for the pressure and the shock speed stated at prop. 21.

- prop. 23, subcase (c) (if ignoring the contact wave with infinite speed). We can also view subcase (a) as a particular case where the velocities of (C) and (UC) are equal

(in which case $\bar{p} = 0$ in the cluster and the interface moves with the common velocity). Again, these two cases are consistent with (IV.24) and (IV.25).

Interface (UC)-(V): it appears in

- prop. 21, subcase (b) as two contact waves between the unclustered left and right states and the vacuum intermediate state,
- prop. 23, subcase (b) where the first wave is a contact between the left unclustered state and the vacuum middle state.

In both cases, the velocity of the interface is that of the non-vacuum states and, of course, the pressure is identically zero. Therefore, the situation is as depicted in Formal Statement 1.

Interface (C)-(V): it appears in

- prop. 23, subcase (b), where the second wave is a contact wave between the vacuum middle state and the right clustered state. We notice that in this case, the clustered state must have zero pressure (otherwise, a declustering wave instantaneously relaxes the pressure to zero),
- prop. 24, subcase (c), where the left and right clustered states are seperated by a vacuum intermediate state. Again, in this case, the pressure inside the clusters is identically zero.

In both cases, the velocity of the (C)-(V) interface is that of the cluster. Therefore, the situation is as depicted in the Formal Statement 1.

Interface (UC)-(UC): it appears in

- prop. 21, subcase (a). We see that this situation is that of a standard contact discontinuity for the uncongested system. The velocities on the uncongested states are equal and equal to that of the interface, and of course, the pressure is identically zero. Therefore, the situation is again as depicted is the Formal Statement 1.

We feel that these observation provide a very strong support to the Formal Statement 1. As pointed out above, this statement allows to close the system (IV.8)-(IV.11) at least until clusters meet. In the one-dimensional framework, proposition 14 provides the cluster collision dynamics. The investigation of cluster dynamics in the two-dimensional case is still work in progress.

4 Conclusion

In this paper, we have studied a continuum model describing a system of particles with short-range repulsive and long-range attractive interaction. This is a model for the study of gregariousness among mammal species for instance. We have focused on the effect of the short-range repulsion and looked at the regime where the interaction is turned on suddenly when the local density becomes close to some limit associated to congestion. We have modeled this effect by introducing a perturbation parameter ε and studied the limit $\varepsilon \to 0$. We have shown that, in the limit regime, the congested regions are domains where the flow is incompressible. The complete determination of the limit system requires the

knowledge of the interface conditions at the boundaries of the congested regions. We have derived these conditions by looking at a model one-dimensional situation (corresponding to the normal direction to the interface) and analyzing the solutions of the Riemann problem for the perturbation model (with finite ε). Taking the limit $\varepsilon \to 0$ in the solutions of the Riemann problem allowed us to provide the missing conditions at the interfaces.

In future work, we will continue the investigation of the interface conditions at the cluster interfaces and will try to show their relevance beyond the special case of one-dimensional Riemann problem solutions. We will also further investigate cluster collision dynamics, in dimensions more than one, since this analysis is not accessible by Riemann problem techniques. At the numerical level, we will seek numerical methods for the resolution of the constrained hyperbolic problem and we will perform numerical comparisons between the particle and continuum models. Two numerical schemes

A Derivation of a macroscopic model of short-range repulsive and long-range attractive interactions

A.1 Individual Based Model with speed and congestion constraints.

We consider N particles in $\mathbb{R}^2$ labeled by $k \in \{1, .., N\}$. These particles are discs of radii d. The motion of the particles is described by the time evolution of their positions X_k and velocity vectors ω_k. Like in the Vicsek algorithm [16, 22, 47], the velocity magnitude of each particle is the same, is constant in time and supposed equal to $c > 0$. The velocity direction ω_k belongs to the unity circle $\mathbb{S}^1 = \left\{\omega \in \mathbb{R}^2, |\omega|^2 = 1\right\}$. This is a usual assumption in the modeling of several biological systems like flocks of birds [3], schools of fish [23, 25] or herds of sheep [39, 40].

We start with a simple continuous-in-time model of a particle system subject to attractive-repulsive binary interactions which describe the aggregation of particles with occupation constraints. The evolution of the positions and velocities is given by:

$$\frac{dX_k}{dt} = c\omega_k, \tag{IV.36}$$

$$\frac{d\omega_k}{dt} = (\mathrm{Id} - \omega_k \otimes \omega_k)(\nu_k^a \xi_k^a - \nu_k^r \xi_k^r), \tag{IV.37}$$

where $\nu_k^a \xi^a$ and $\nu_k^r \xi^r$ are the attractive and repulsive forces respectively. The matrix $(\mathrm{Id} - \omega_k \otimes \omega_k)$ is the orthogonal projector onto the plane orthogonal to ω_k. It is applied to both forces in order to keep the magnitude of the speed constant in time. ξ^a and ξ^r are the local centers of mass of the particle distribution inside interaction discs centered at X_k with radii respectively equal to R_a and R_r

$$\xi_k^a = \frac{\sum\limits_{j,|X_j - X_k| \le R_a} (X_j - X_k)}{\sum\limits_{j,|X_j - X_k| \le R_a} 1}, \quad \xi_k^r = \frac{\sum\limits_{j,|X_j - X_k| \le R_r} (X_j - X_k)}{\sum\limits_{j,|X_j - X_k| \le R_r} 1}.$$

ν_k^a and ν_k^r are scaling factors which provide the intensities of the forces. The repulsive force radius R_r is supposed much smaller than the attractive force radius R_a. The resulting force attracts the particles towards the center of mass of the particle distribution at large distances and repels them from the center of mass of the particle distribution at short distances. To some extent, it is an implementation of the attractive-repulsive scheme proposed by Couzin [16].

Finally, we suppose that ν_a is a constant and ν_r depends on the local density inside the repulsive interaction disc:

$$\nu_k^a = \nu_a, \quad \nu_k^r = \nu_r\left(\rho_k^r\right), \quad \rho_k^r = \frac{\pi d^2 \sum_{j,|X_j - X_k| \leq R_r} 1}{\pi R_r^2},$$

where ν_r is an increasing function. The function ν_r prevents the local density from exceeding the maximal density ρ^* which corresponds to the case where all particles are in contact with their neighbours. Clearly, ρ^* is the ratio of the maximal occupied surface in a disk of radius R_r by disks of radii d and is of the order of unity. Therefore, the function ν_r tends to infinity as $\rho_k^r \to \rho^*$. We defer the explicit choice of the function ν_r to the end of the section.

A.2 Mean-field model, hydrodynamic limit and macroscopic model

The goal of this appendix is to provide a model for large systems of interacting particles according to (IV.36)-(IV.37) at large time and space scales. For this purpose, we will perform a sequence of rescalings. The first rescaling aims at taking into account the large number of interacting particles: it leads to the so-called mean field model. Assuming that the system is included in a fixed box, the limit $N \to +\infty$ implies that the area πd^2 occupied by each particle tends to 0 like $1/N$ in such a way that the total area $N\pi d^2$ occupied by the particles remains constant. We denote by $\alpha = \lim_{N\to+\infty} N\pi d^2$ the fraction of the surface occupied by the particles. The second rescaling is a hydrodynamic scaling where large time and space scales are considered. Both scaling are classical and are well detailed in [13, 20]. They have been applied to swarming model in [11, 22, 27].

A.2.1 Formal derivation of the mean-field model

We refer to [45] for classical references on the mean-field limit. We consider the empirical distribution $f^N(x, \omega, t)$ defined by

$$f^N(x, \omega, t) = \frac{1}{N} \sum_{k=1}^{N} \delta(x - X_k(t))\delta(\omega, \omega_k(t)).$$

$\delta(x)$ denote the Dirac delta on $\mathbb{R}^2$, while $\delta(\omega, \omega_0)$ denotes the Dirac delta on $\mathbb{S}^1$ centered at ω_0 (i.e. $\delta(\omega, \omega_0)$ is the probability measure supported by $\{\omega_0\}$). It is an easy matter to check that f^N satisfies the following kinetic equation

$$\partial_t f^N + c\omega \cdot \nabla_x f^N + \nabla_\omega \cdot \left(\left(F_a^N - F_r^N\right) f^N\right) = 0,$$

where F_a^N and F_r^N are the attractive and repulsive forces, given by

$$F_a^N(x,\omega,t) = \nu_a(\mathrm{Id} - \omega\otimes\omega)\xi_a^N, \qquad F_r^N(x,\omega,t) = \nu_r^N(\mathrm{Id} - \omega\otimes\omega)\xi_r^N,$$

with

$$\xi_a^N(x,\omega,t) = \frac{\int K_a(y-x)(y-x)\rho^N(y,t)dy}{\int K_a(y-x)\rho^N(y,t)dy},\ \xi_r^N(x,\omega,t) = \frac{\int K_r(y-x)(y-x)\rho^N(y,t)dy}{\int K_r(y-x)\rho^N(y,t)dy},$$
$$\nu_r^N = \nu_r\left(\frac{N\pi d^2\int K_r(y-x)\rho^N(y,t)dy}{\int K_r(y-x)(y,t)dy}\right),$$

where $\rho^N(x,t) = \int_{\omega\in\mathbb{S}^1} f^N(x,\omega,t)d\omega$ is the local density and K_a (resp. K_r) is the indicator function of the disc of radius R_a (resp. R_r). Here, it is clear that more general kernels K_a, K_r can be used.

The formal mean-field limit $N\to+\infty$ of this model is (we recall that $\alpha = \lim_{N\to+\infty} N\pi d^2$):

$$\partial_t f + c\omega\cdot\nabla_x f + \nabla_\omega\cdot((F_a - F_r)\,f) = 0, \tag{IV.38}$$
$$F_a(x,\omega,t) = \nu_a(\mathrm{Id} - \omega\otimes\omega)\xi_a, \qquad F_r(x,\omega,t) = \nu_r(\mathrm{Id} - \omega\otimes\omega)\xi_r, \tag{IV.39}$$
$$\xi_a(x,\omega,t) = \frac{\int K_a(y-x)(y-x)\rho(y,t)dy}{\int K_a(y-x)\rho(y,t)dy}, \quad \xi_r(x,\omega,t) = \frac{\int K_r(y-x)(y-x)\rho(y,t)dy}{\int K_r(y-x)\rho(y,t)dy} \tag{IV.40}$$
$$\nu_r = \nu_r\left(\frac{\alpha\int K_r(y-x)\rho(y,t)dy}{\int K_r(y-x)dy}\right), \rho(x,t) = \int f(x,\omega,t)d\omega. \tag{IV.41}$$

Rigorous justifications of this limit are outside the scope of this article.

A.2.2 Hydrodynamic scaling

In order to select the relevant scales, we first rewrite our system in dimensionless variables. We consider a space scale x_0 (typically the range R_r of the repulsive force) and we choose a time scale $t_0 = x_0/c$. The associated dimensionless time and space variables are $t' = t/t_0$ and $x' = x/x_0$. We also introduce scaled collision kernels $K'_{a,r}$ such that $K_{a,r}(x_0x') = K'_{a,r}(x)$, scaled intensities $\nu'_a = \nu_a x_0^2/c$, $\nu'_r = \nu_r x_0^2/c$ and a scaled distribution function $f' = \alpha f$. In the case where $K_{a,r}$ are indicator functions of balls of radii $R_{a,r}$, this amounts to rescaling the radii to new values $R'_{a,r} = R_{a,r}/x_0$. After removing the primes, the system in the new variables and with the new unknowns is similar to (IV.38)-(IV.41) but with $c = 1$ and $\alpha = 1$.

To derive the large time and space dynamics, we introduce the following change of variables $\tilde{x} = \eta x$, $\tilde{t} = \eta t$ with $\eta \ll 1$. In the new variables, the distribution function $f^\eta(\tilde{x},\omega,\tilde{t}) = f(x,\omega,t)$ satisfies the following system (omitting the tildes):

$$\eta\left(\partial_t f^\eta + \omega\cdot\nabla_x f^\eta\right) + \nabla_\omega\cdot((F_a^\eta - F_r^\eta)\,f^\eta) = 0,$$
$$F_a^\eta(x,\omega,t) = \nu_a^\eta(\mathrm{Id} - \omega\otimes\omega)\xi_a^\eta,\ \xi_a^\eta(x,\omega,t) = \frac{1}{\eta}\frac{\int K_a^\eta\left(\frac{y-x}{\eta}\right)(y-x)\rho^\eta(y,t)dy}{\int K_a^\eta\left(\frac{y-x}{\eta}\right)\rho^\eta(y,t)dy},$$
$$F_r^\eta(x,\omega,t) = \nu_r^\eta(\mathrm{Id} - \omega\otimes\omega)\xi_r^\eta,\ \xi_r^\eta(x,\omega,t) = \frac{1}{\eta}\frac{\int K_r^\eta\left(\frac{y-x}{\eta}\right)(y-x)\rho^\eta(y,t)dy}{\int K_r^\eta\left(\frac{y-x}{\eta}\right)\rho^\eta(y,t)dy},$$
$$\nu_r^\eta = \nu_r^\eta\left(\frac{\frac{1}{\eta^2}\int K_r^\eta\left(\frac{y-x}{\eta}\right)\rho^\eta(y,t)dy}{\frac{1}{\eta^2}\int K_r^\eta\left(\frac{y-x}{\eta}\right)dy}\right),\ \rho^\eta(x,t) = \int f^\eta(x,\omega,t)d\omega,$$

where K_a^η and K_r^η are the scaled interaction kernels and ν_a^η, ν_r^η, the scaled intensities.

We first suppose that the repulsive kernel K_r^η and the repulsive intensity ν_r^η are unchanged in the scaling: $K_r^\eta = K_r$, $\nu_r^\eta(\rho) = \nu_r(\rho)$. This means that the range of the repulsive force is supposed of order η. To analyze the limit $\eta \to 0$, we first need an expansion of ξ_r^η in terms of η. The following lemma provides the result for an isotropic kernel K_r ($K_r(z) = K_r(|z|)$)

Lemma 25. *Under suitable regularity assumptions on ρ^η, we have the expansion*

$$\begin{aligned}
&\frac{1}{\eta^2}\int K_r\left(\left|\frac{y-x}{\eta}\right|\right)\rho^\eta(y)dy = a\rho^\eta(x) + o(\eta),\\
&\xi_r^\eta(x,\omega,t) = \eta\frac{B\nabla_x\rho^\eta(x)}{a\rho^\eta(x)} + o(\eta),\\
&\nu_r^\eta\left(\frac{\frac{1}{\eta^2}\int K_r^\eta\left(\frac{y-x}{\eta}\right)\rho^\eta(y,t)dy}{\frac{1}{\eta^2}\int K_r^\eta\left(\frac{y-x}{\eta}\right)dy}\right) = \nu_r(\rho) + o(1),
\end{aligned}$$

where $a = \int K_r(|z|)dz$ and $B = \int K_r(|z|)z\otimes z dz = (\int K_r(r)r^3dr\, Id)$.

The proof of this lemma is elementary and omitted. In the case where K_r is the indicator function of the disc of radius R_r, the coefficients a and B are equal to $a = \pi R_r^2$ and $B = b\mathrm{Id}$, where $b = \pi\frac{R_r^4}{4}\mathrm{Id}$. Now we consider the scaling of the attractive kernel K_a^η and attractive intensity ν_a^η. We suppose that the attractive force remains non-local as η tends to 0 and weaker than the repulsive force. To express these assumptions, we suppose that the scaled attractive kernel K_a^η and intensity ν_a^η are given by

$$K_a^\eta(z) = K_a(\eta z), \quad \nu_a^\eta = \eta^2\nu_a.$$

For simplicity, we choose $\nu_a = 1$. We also absorb the ratio b/a into the definition of ν_r. In particular, in the case where K_r is the indicator of the ball of radius R_r, we can fix $x_0 = R_r/2$.

Under all these modelling assumptions and thanks to lemma 25, the system can be written formally, in the limit $\eta \to 0$:

$$\partial_t f + \omega\cdot\nabla_x f + \nabla_\omega\cdot((F_a - F_r)\,f) = 0, \tag{IV.42}$$

$$F_a(x,\omega,t) = (\mathrm{Id}-\omega\otimes\omega)\xi_a,\ \xi_a(x,t) = \left(\frac{\int K_a(|y-x|)\,(y-x)\rho(y,t)dy}{\int K_a(|y-x|)\,\rho(y,t)dy}\right) \tag{IV.43}$$

$$F_r(x,\omega,t) = \nu_r(\rho)(\mathrm{Id}-\omega\otimes\omega)\widetilde{\xi}_r,\ \widetilde{\xi}_r(x,t) = \frac{\nabla_x\rho(x,t)}{\rho(x,t)}. \tag{IV.44}$$

A.2.3 Macroscopic model

The last step is to obtain the dynamics of macroscopic quantities associated to the flow. Here we will only consider the density and momentum. We find that under suitable regularity and decay assumptions on f, the density $\rho = \int f d\omega$ and momentum $\rho\Omega =$

$\int f\omega d\omega$ satisfy the following system of mass and momentum balance equations:

$$\partial_t \rho + \nabla_x \cdot \rho\Omega = 0, \tag{IV.45}$$
$$\partial_t \rho\Omega + \nabla_x \cdot \left(\int f\omega \otimes \omega d\omega \right) = \left(\int (\mathrm{Id} - \omega \otimes \omega) f d\omega \right) \left(\xi_a - \nu_r(\rho)\widetilde{\xi}_r \right), \tag{IV.46}$$
$$\xi_a(x,t) = \left(\frac{\int K_a\left(|y-x|\right)(y-x)\rho(y,t)dy}{\int K_a\left(|y-x|\right)\rho(y,t)dy} \right), \quad \widetilde{\xi}_r(x,t) = \frac{\nabla_x \rho(x,t)}{\rho(x,t)}. \tag{IV.47}$$

To close system (IV.45)-(IV.46), we assume that f is a monokinetic distribution:

$$f(x,\omega,t) = \rho(x,t)\delta(\omega, \Omega(x,t)),$$

with $|\Omega(x,t)| = 1$. This assumptions presupposes that a local equilibrium is reached where all particles are locally aligned. Although no justification of this assumption can be made at this point, the features displayed by the system seem meaningful in view of gregariousness modelling. We find:

$$\partial_t \rho + \nabla_x \cdot \rho\Omega = 0, \tag{IV.48}$$
$$\partial_t \left(\rho\Omega\right) + \nabla_x \cdot \left(\rho\Omega \otimes \Omega\right) = \rho(\mathrm{Id} - \Omega \otimes \Omega)(\xi_a - \nu_r(\rho)\widetilde{\xi}_r), \tag{IV.49}$$

where ξ_a and $\widetilde{\xi}_r$ are given by (IV.47). Factoring out ρ in (IV.49), using (IV.48), we also get the following form of the system:

$$\begin{aligned}
&\partial_t \rho + \nabla_x \cdot \rho\Omega = 0,\\
&\partial_t \Omega + \Omega \cdot \nabla_x \Omega + \nu_r(\rho)(\mathrm{Id} - \Omega \otimes \Omega)\widetilde{\xi}_r = (\mathrm{Id} - \Omega \otimes \Omega)\xi_a,\\
&\xi_a(x,t) = \left(\frac{\int K_a\left(|y-x|\right)(y-x)\rho(y,t)dy}{\int K_a\left(|y-x|\right)\rho(y,t)dy} \right), \quad \widetilde{\xi}_r(x,t) = \frac{\nabla_x \rho(x,t)}{\rho(x,t)}.
\end{aligned}$$

A.3 Repulsive force intensity and macroscopic model

Let us return now to the choice of the function ν_r. This function tends to infinity when $\rho \to \rho^*$. Like in the traffic model devised in [6], we assume that this function behaves like ρ^γ when $\rho \ll \rho^*$ and tends to infinity when $\rho \to \rho^*$. The prototype of such a function is

$$p(\rho) = \frac{1}{\left(\frac{1}{\rho^*} - \frac{1}{\rho}\right)^\gamma}, \tag{IV.50}$$

where $\gamma \geq 1$. We will keep this example constantly in the paper for simplicity but the results are valid for all functions having the same properties. We consider that $\nu_r(\rho) = \rho p'(\rho)$. In this way, we suppose that repulsion acts like a standard presure force in a gas, but, when the density reaches the congestion density ρ^*, the pressure tends to infinity. Since the equation for Ω is used instead of that for $\rho\Omega$, the interpretation of p in standard gas dynamics terms would rather be that of an enthalpy (i.e. $p'(\rho) = P'(\rho)/\rho$ where P is the actual fluid mechanical pressure), but the results would be similar if we considered the equation for $\rho\Omega$ instead. Indeed, because of the constraint $|\Omega| = 1$, the system is non-conservative in the projection term $(\mathrm{Id} - \Omega \otimes \Omega)$. Finally, we get the

following system

$$\partial_t \rho + \nabla_x \cdot \rho\Omega = 0, \tag{IV.51}$$
$$\partial_t \Omega + \Omega \cdot \nabla_x \Omega + (\mathrm{Id} - \Omega \otimes \Omega)\nabla_x p(\rho) = (\mathrm{Id} - \Omega \otimes \Omega)\xi_a, \tag{IV.52}$$
$$\xi_a(x,t) = \left(\frac{\int K_a\left(|y-x|\right)(y-x)\rho(y,t)dy}{\int K_a\left(|y-x|\right)\rho(y,t)dy} \right). \tag{IV.53}$$

This system provides the starting point of the present article. Since this paper is focused on the treatment of congestion phenomena, we remove the non-local attractive force. Indeed, this term is a zero-th order derivative term and does not intervene in the jump relations across discontinuities.

B Conservative laws for the one-dimensional system

In this appendix, we are looking for conservative forms of the one-dimensional system (IV.20)-(IV.21). The most general conservative form is written:

$$\partial_t g(\rho,\theta) + \partial_x f(\rho,\theta) = 0. \tag{IV.54}$$

where g and f are smooth functions of ρ and θ. The following proposition exhibits an infinite set of such conservative forms.

Proposition 26. *If (g,f) is a conservative form of (IV.20)-(IV.21), then their partial derivatives are related by*

$$\frac{\partial f}{\partial \rho} = \frac{\partial g}{\partial \rho}\cos\theta - \frac{\partial g}{\partial \theta}\sin\theta p'(\rho), \quad \frac{\partial f}{\partial \theta} = \frac{\partial g}{\partial \theta}\cos\theta - \frac{\partial g}{\partial \rho}\rho\sin\theta. \tag{IV.55}$$

Moreover, if g is a function with separated variables $g(\rho,\theta) = u(\theta)v(\rho)$, then u and v satisfy

$$\rho v''(\rho) = kp'(\rho)v(\rho), \tag{IV.56}$$
$$u''(\theta) + (cotan\theta)u'(\theta) = ku(\theta), \tag{IV.57}$$

where k is a constant real number. Each $k \in \mathbb{R}$ gives rise to possible (g,f) pairs.

Proof. Performing the chain rule in (IV.54) and using (IV.20),(IV.21), we easily get (IV.55). Then, using that differentiations with respect to ρ and θ commute, (IV.55) gives rise to an elliptic equation satisfied by g and inserting the hypothesis of separated variables, we obtain (IV.56),(IV.57).

Equation (IV.57) is the Legendre differential equation (in polar coordinates). The two-dimensional vector space of solutions of this equation is spanned by the Legendre functions of first and second species and each of them gives rise to possible (g,f) pairs. ∎

The solutions of (IV.56) exists for all $k \in \mathbb{R}$. However, they have a priori no explicit expression exept for $k = 0$. In this case, the 2-dimensional vector space of solutions of

(IV.56) is spanned by $\{1, \rho\}$ and for (IV.57), is $\{1, \Psi(\cos\theta)\}$. We can actually check that the following (g, f) pairs

$$(g, f) = (\rho, \rho\cos\theta), \tag{IV.58}$$
$$(g, f) = (\Psi(\cos\theta), \Phi(\cos\theta) + p(\rho)), \tag{IV.59}$$
$$(g, f) = (\rho\Psi(\cos\theta), \rho\cos\theta\Psi(\cos\theta) + P(\rho)), \tag{IV.60}$$

where P is an antiderivative of $\rho p'(\rho)$, are non trivial solutions. The conservative form studied in this article corresponds to the pairs (IV.58) and (IV.59). The pairs (IV.58) and (IV.60) form another such conservative system.

C Proof of proposition 14 (cluster collisions)

Proof. 1- Let $x_0 \in [a(t_c), m]$ and $h \in \mathcal{C}_c^\infty(D'')$, where D'' is a neighbourhood of x_0 in D (cf. figure IV.5). Applying the Green's formula on the domain D'', we obtain:

$$< \partial_t\Psi(\cos\theta) + \partial_x\Phi(\cos\theta), h >= (\Psi(\cos\theta) - \Psi(\cos\theta_\ell)) \int_{x_0-\eta}^{x_0+\eta} h(t_c, x)dx, \tag{IV.61}$$

where $< ., . >$ denotes the duality brackets. Since we look for $\bar{p}(x, t) = \pi(x)\delta(t - t_c)$, we also have

$$- < \partial_x\bar{p}, h >= -\int_{x_0-\eta}^{x_0+\eta} \partial_x\pi(x)h(t_c, x)dx. \tag{IV.62}$$

If (IV.23) is satisfied, then equations (IV.61) and (IV.62) imply that

$$(\Psi(\cos\theta) - \Psi(\cos(\theta_\ell))) = -\partial_x\pi(x_0).$$

The same arguments (for any $x \in [a(t_c), b(t_c)]$) lead to

$$-\partial_x\pi(x) = \begin{cases} (\Psi(\cos\theta) - \Psi(\cos(\theta_\ell))), & \text{if } x \in [a(t_c), m], \\ (\Psi(\cos\theta) - \Psi(\cos(\theta_r))), & \text{if } x \in [m, b(t_c)], \end{cases}$$

and (supposing π continuous) to

$$\pi(x) = \begin{cases} (\Psi(\cos\theta) - \Psi(\cos(\theta_\ell)))(m - x) & \\ \quad +(\Psi(\cos\theta) - \Psi(\cos(\theta_r)))(b(t_c) - m), & \text{if } x \in [a(t_c), m], \\ (\Psi(\cos\theta) - \Psi(\cos(\theta_r)))(b(t_c) - x), & \text{if } x \in [m, b(t_c)], \end{cases}$$

Supposing that $\bar{p}$ and then π equal zero outsidse the clusters, we get

$$\pi(a(t_c)) = (\Psi(\cos\theta) - \Psi(\cos(\theta_\ell)))(m - a(t_c)) + (\Psi(\cos\theta) - \Psi(\cos(\theta_r)))(b(t_c) - m) = 0.$$

2 - Let h be a test function in the neighbourhood D' of D (cf. figure IV.5). We denote $D_1 = D \cap \{t \le t_c\}$ and $D_2 = D \cap \{t \ge t_c\}$. Applying Green's formula, we easily obtain:

$$< \partial_t\rho + \partial_x(\rho\cos\theta), h >= -\rho^* \int_{a_\ell(t_c)}^{b_\ell(t_c)} h(t_c, x)dx - \rho^* \int_{a_r(t_c)}^{b_r(t_c)} h(t_c, x)dx + \rho^* \int_{a(t_c)}^{b(t_c)} h(t_c, x)dx$$

Thus, the density equation (IV.22) (satisfied in the distributional sense) stems from the identity

$$(b_\ell(t_c) - a_\ell(t_c)) + (b_r(t_c) - a_r(t_c)) = (b(t_c) - a(t_c)).$$

If we now apply Green's formula with a test function $h \in \mathcal{C}_c^\infty(D)$, we easily show that eq. (IV.23) is satisfied in the distributional sense. Note that in this case, the test function has a compact support in D since $\Psi(\theta)$ is not defined in the vacuum region $\rho = 0$. ∎

D Proof of proposition 16 (study of the Hugoniot loci)

In this section, we provide a detailed study of the Hugoniot curves. Let $(\rho_\ell, \theta_\ell) \in]0, \rho^*[\times]0, \pi[$ be an arbitrary left state. We need to find the geometric behaviour of the Hugoniot loci associated to this left state. The classical theory of nonlinear conservation laws provides only information on the local behaviour of $\mathcal{H}^\varepsilon_\pm$. Each $\mathcal{H}^\varepsilon_+$, $\mathcal{H}^\varepsilon_-$ consists of a one-dimensional manifold tangent to the integral curves of the right eigenvectors up to the second order. In the $(\rho, \Psi(\cos\theta))$-plane, the 1-Hugoniot curve $\mathcal{H}^\varepsilon_-$ is thus locally decreasing and the 2-Hugoniot curve $\mathcal{H}^\varepsilon_+$ is locally increasing because of the direction of the vectors $r^\varepsilon_\pm = (\pm\rho\,|\sin\theta|\,, \sqrt{\varepsilon p'(\rho)\rho})$. In the (ρ, θ)-plane, the 1-Hugoniot curve $\mathcal{H}^\varepsilon_-$ defines a locally increasing function $\theta = (h^\varepsilon_-)^{-1}(\rho)$ while the 2-Hugoniot curve $\mathcal{H}^\varepsilon_+$ defines a locally decreasing function $\theta = (h^\varepsilon_+)^{-1}(\rho)$. Actually, this property is global (i.e. $(h^\varepsilon_-)^{-1}$ (resp. $(h^\varepsilon_+)^{-1}$) is a globally increasing (resp. decreasing) function of ρ for all $\rho \in]0, \rho^*[$). To prove this, let us begin with a simple and useful lemma.

Lemma 27. *For all $u \in [-1, 1]$, the function $f_u : v \in]-1, 1[\to \Phi(v) - u\Psi(v)$ is convex and has a minimum at the point u. In particular, we have*

$$\forall v \neq u, \quad (\Phi(v) - \Phi(u)) - u\,(\Psi(v) - \Psi(u)) = f_u(v) - f_u(u) > 0.$$

The proof is elementary and omitted. We now analyze the behaviour of $\mathcal{H}^\varepsilon_\pm$ in more detail when ε becomes small. Proposition 16 is an immediate consequence of the following lemma.

Lemma 28. *The behaviour of $\mathcal{H}^\varepsilon_+$, $\mathcal{H}^\varepsilon_-$ does not depend on the left state. Let (ρ_ℓ, θ_ℓ) be a left state. Then:*

(i) Suppose θ_r is fixed. The function $\rho_r \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho_r, \theta_r)$ has at most two zeros and there exists $\varepsilon' > 0$ such that for all $\varepsilon < \varepsilon'$, the function $\rho_r \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho_r, \theta_r)$ has only one positive zero. This zero tends to ρ^* as ε tends to 0.

(ii) Suppose ρ_r is fixed. Then $\forall \varepsilon > 0$, the function $\theta_r \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho_r, \theta_r)$ has two zeros, one lower and one larger than θ_ℓ, and both of them tend to θ_ℓ as ε tends to 0.

The Hugoniot locus tends to the union of the straight lines $\{\theta = \theta_\ell\}$ and $\{\rho = \rho^*\}$.

Note that these results imply that the Hugoniot locus consists of two monotonous curves as functions of ρ (otherwise H with fixed θ_r would have more than two zeros). The local behaviour of the Hugoniot locus enables us to determine that the increasing curve is associated to the first eigenvalue λ^ε_- and the decreasing curve to the second eigenvalue λ^ε_+.

Proof. (i) Let us fix the left state (ρ_ℓ, θ_ℓ) and the right angle θ_r. So as to get a more readible proof, the function $\rho_r \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho_r, \theta_r)$ will be denoted by H but its derivative will be denoted by a partial derivative $\partial_{\rho_r} H$. We look for the zero set of H in the interval $]0, \rho^*[$. As it can be easily checked, the function H is strictly convex and thus has at most two zeros. Moreover, the value of H at $\rho_r = \rho_\ell$ is strictly negative, if θ_r is not equal to

θ_ℓ. Like the function p, H tends to $+\infty$ when ρ_r tends to the maximal density ρ^*. Then H has only one zero in $]\rho_\ell, \rho^*[$. We have

$$H(0) = \rho_\ell \left(\varepsilon p(\rho_\ell) + [\Psi(\cos\theta)]\cos\theta_\ell - [\Phi(\cos\theta)]\right).$$

Lemma 27 implies that the second term of this expression is strictly negative and thus $H(0)$ becomes stricly negative for small ε. Thanks to its convexity, we deduce that there exists ε' such that for all $\varepsilon < \varepsilon'$, the function H has no zero in the interval $]0, \rho_\ell[$.

To show that the only zero of H tends to ρ^*, let us rewrite H as follows

$$H(\rho_r) = [\varepsilon p(\rho)]\,[\rho] + [\Phi(\cos\theta)]\,[\rho] - [\Psi(\cos\theta)]\,[\rho]\cos\theta_r - [\Psi(\cos\theta)]\,[\cos\theta]\,\rho_\ell,$$

and thanks to lemma 27, the zero of H satisfies

$$[\varepsilon p(\rho)]\,[\rho] \geq [\Psi(\cos\theta)]\,[\cos\theta]\,\rho_\ell > 0.$$

So we can easily conclude that the zero of H tends to ρ^*.

(ii) Like in the first point, let us denote the function $\theta_r \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho_r, .)$ by H. First, the value taken by H at $\theta_r = \theta_\ell$ is positive. Some easy computations leads to the following expression of the first (partial) derivative of H:

$$\frac{\partial H}{\partial \theta_r}(\theta_r) = \frac{1}{\sin\theta_r}\left(\rho_\ell\,[\cos\theta] + [\Psi(\cos\theta)]\,\rho_r \sin^2\theta_r\right).$$

As ρ_ℓ and ρ_r are positive, the sign of the derivative is the same as the sign of $[\cos\theta]$. Thus H is increasing on $[0, \theta_\ell]$ and decreasing on $[\theta_\ell, \pi]$. Moreover using the fact that $\Psi(u) = \Phi(u) + \log(1+u)$, we can write H as

$$H(\theta_r) = \Phi(\cos\theta_r)\,[\rho(1-\cos\theta)] + \log(1+\cos\theta_r)\,[\rho\cos\theta] + A(\varepsilon, \rho_r, \theta_r),$$

where A is a bounded function. It implies that H tends to $-\infty$ when θ_r tends to 0. In the same way and by using the identity $\Psi(u) = -\Phi(u) + \log(1-u)$, we can show that H tends also to $-\infty$ when θ_r tends to π. We deduce that H has exactly two zeros.

Let us remark that

$$H_\varepsilon = H_1 - (1-\varepsilon)\,[p(\rho)]\,[\rho].$$

This implies that $H_\varepsilon^{-1}(0) = H_1^{-1}((1-\varepsilon)\,[p(\rho)]\,[\rho])$ and then that the zeros of H_ε tend to θ_ℓ as ε tends to 0. ∎

E Proof of proposition 17 (study of the integral curves of the right eigenvectors)

Proof. 1. We easily check that $\theta'(\rho) = \mp\sqrt{\varepsilon p'(\rho)/\rho}$, leading to the result.

2. For a fixed ρ, the quantity

$$\theta^\varepsilon = (i_\pm^\varepsilon)^{-1}(\rho) = \theta_\ell \mp \sqrt{\varepsilon}\left(\int_{\rho_\ell}^{\rho} \sqrt{\frac{\varepsilon p'(u)}{u}}du\right)$$

converges to θ_ℓ as ε goes to 0. For a fixed θ, the quantity $\sqrt{\varepsilon}\left(\int_{\rho_\ell}^{\rho^\varepsilon}\sqrt{\frac{\varepsilon p'(u)}{u}}du\right) = \theta - \theta_\ell$ is a constant. So as ε tends to 0, the integral term has to tend to $+\infty$, which implies the convergence of $\rho^\varepsilon = i_+^\varepsilon(\theta)$ to ρ^*. Besides, the function inside the integral behaves like $O\left(\sqrt{\varepsilon}(\rho^* - u)^{-\frac{\gamma+1}{2}}\right)$ when $\rho \to \rho^*$. This leads to a diverging integral for $\gamma > 1$. Then the integral behaves like $O\left(\sqrt{\varepsilon}(\rho^* - \rho_d^\varepsilon)^{-\frac{\gamma-1}{2}}\right)$ and thus we get $\rho^* - i_\pm^\varepsilon(\theta) = O\left(\varepsilon^{\frac{1}{\gamma-1}}\right)$.

3. Let $\varepsilon' > 0$ and $\rho < \rho_r^{\varepsilon'}$. From the rarefaction curve equation (IV.34), $(i_\pm^\varepsilon)^{-1}(\rho)$ satisfies

$$|(i_\pm^\varepsilon)^{-1}(\rho) - \theta_r| \leq \int_0^{\rho_r^\varepsilon}\sqrt{\frac{\varepsilon p'(u)}{u}}du = \sqrt{\varepsilon}\int_0^{\rho_r^\varepsilon}\frac{\gamma u^{\frac{\gamma-2}{2}}\rho^{*\gamma+1}}{(\rho^* - u)^{\frac{\gamma+1}{2}}}du.$$

Assuming that the limit of $\varepsilon p(\rho_r^\varepsilon)$ is finite, we get $\rho^* - \rho_r^\varepsilon = O(\varepsilon^{\frac{1}{\gamma}})$. Thus, the function inside the integral behaves like $\sqrt{\varepsilon}(\rho^* - u)^{-\frac{\gamma+1}{2}}$ when $\rho \to \rho^*$. This leads to a diverging integral for $\gamma > 1$, and then the integral behaves like $O\left(\sqrt{\varepsilon}(\rho^* - \rho_r^\varepsilon)^{-\frac{\gamma-1}{2}}\right)$ and thus like $O(\varepsilon^{\frac{1}{2\gamma}})$. ∎

F Proofs of theorem 19 and proposition 20 (solutions of the Riemann problem for $\varepsilon > 0$)

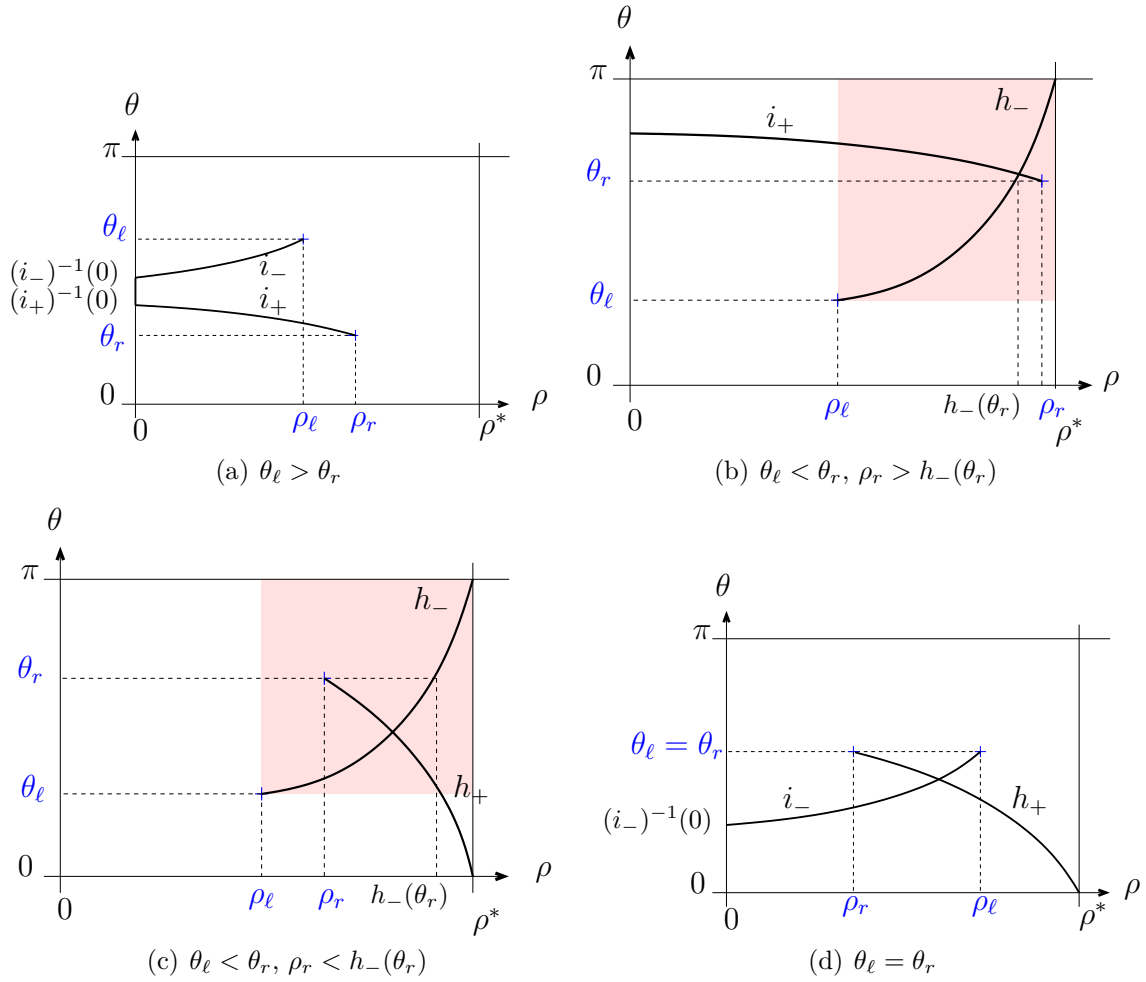
F.1 Proof of theorem 19

Proof. Let (ρ_ℓ, θ_ℓ) and (ρ_r, θ_r) be left and right states respectively and let us suppose that the intersection of the 1-forward wave curve $W_-^{f,\varepsilon}$ issued from the left state and the 2-backward wave curve $W_+^{b,\varepsilon}$ issued from the right state reduces to one point $(\widetilde{\rho}, \widetilde{\theta})$ (in all the proof, the 1-wave will be implicitly relative to the left state while the 2-wave curve will be implicitly relative to the right state). To find where the intersection on the curves is located, the main arguments will be the monotony of the wave curves given by propositions 16 and 17 (independently of the location of the states in the (ρ, θ)-plane) and their convergence speed to their asymptotic limit. The wave curves will be considered as functions of θ in their domain of definition:

$$w_-^\varepsilon = \begin{cases} i_-^\varepsilon & \text{for } \theta \in [(i_-^\varepsilon)^{-1}(0), \theta_\ell], \\ h_-^\varepsilon & \text{for } \theta \in [\theta_\ell, \pi[, \end{cases} \qquad w_+^\varepsilon = \begin{cases} h_+^\varepsilon & \text{for } \theta \in [0, \theta_r], \\ i_+^\varepsilon & \text{for } \theta \in [\theta_r, (i_+^\varepsilon)^{-1}(0)[, \end{cases}$$

where w_- is an increasing function and w_+ is a decreasing function. The functions $h_\pm$ and $i_\pm$ are respectively defined in propositions 16 and 17. Let us examine the different cases suggested by the theorem successively: θ_ℓ greater or lower or equal to θ_r. For the reader's convenience, the corresponding geometric configurations of the wave curves are illustrated in figure IV.12.

Case $\theta_\ell > \theta_r$ (Fig. IV.12, (a)). From proposition 17, $(i_-^\varepsilon)^{-1}(0)$ (resp. $(i_+^\varepsilon)^{-1}(0)$) tends to θ_ℓ (resp. θ_r) as ε goes to zero (and the third point of the same proposition asserts that it is still the case when ρ_r tends to ρ^*). So, assuming that there exists α such that $(i_-^\alpha)^{-1}(0) < (i_+^\varepsilon)^{-1}(0)$, there exists $\beta < \alpha$ such that $\theta_r < (i_-^\beta)^{-1}(0) = (i_+^\beta)^{-1}(0) < \theta_\ell$. So, since the domains of definition of w_- and w_+ are respectively $[(i_-^\beta)^{-1}(0), \pi[$ and $]0, (i_+^\beta)^{-1}(0)]$, the only intersection point of w_+^β and w_-^β is the intersection of the 1 and



(a) $\theta_\ell > \theta_r$ (b) $\theta_\ell < \theta_r$, $\rho_r > h_-(\theta_r)$

(c) $\theta_\ell < \theta_r$, $\rho_r < h_-(\theta_r)$ (d) $\theta_\ell = \theta_r$

Figure IV.12: Schematics of the intersections of the wave curves (proof of theorem 19). Only the parts which meet are represented.

2-rarefaction curves at $(i_+^\beta)^{-1}(0)$. As ε decreases, the intersection point disappears since the domains of definition are separated. However, the integral curves meet the $\{\rho = 0\}$ axis at the states $(0, i_-^\varepsilon(0))$, $(0, i_+^\varepsilon(0))$ and these states are connected by vacuum.

Case $\theta_\ell < \theta_r$. We suppose that ρ_ℓ is lower than ρ_r. For all ε, the increasing 1-shock curve issued from the left state divides the domain $[\rho_\ell, \rho^*] \times [\theta_\ell, 2\pi]$ (to which the right state belongs) in two parts: the left domain where the right state is on the left side of the 1-Hugoniot curve issued from the left state ($\rho_r^\varepsilon < h_-^\varepsilon(\theta_r^\varepsilon)$) and the right domain where the right state is on the right side of the 1-Hugoniot curve issued from the left state ($\rho_r^\varepsilon > h_-^\varepsilon(\theta_r^\varepsilon)$).

- Assume that for all ε the right state is on the right side of the 1-shock curve (Fig. IV.12, (b)). We consider $w_-^\varepsilon - w_+^\varepsilon$ on the interval $[(i_-^\varepsilon)^{-1}(0), (i_+^\varepsilon)^{-1}(0)]$, which is the intersection of the domains of definition of w_-^ε and w_+^ε. The function $w_-^\varepsilon - w_+^\varepsilon$ is increasing. We have $(w_-^\varepsilon - w_+^\varepsilon)((i_-^\varepsilon)^{-1}(0)) = -(w_+)((i_-)^{-1}(0)) < 0$ and $(w_-^\varepsilon - w_+^\varepsilon)(\theta_r) = h_-^\varepsilon(\theta_r) - \rho_r > 0$. So the only zero of $w_-^\varepsilon - w_+^\varepsilon$ is in the interval $[(i_-^\varepsilon)^{-1}(0), \theta_r]$. So the intersection point of the two wave curves is the intersection of the 1-shock curve and the 2-rarefaction curve. This corresponds to the second subcase of the third case of the theorem. Note that the

limit of the 1-shock curve (proposition 16) implies that the limit of this case should be considered only if ρ_r tends to ρ^*.

- Assume that for all ε the right state is on the left side of the 1-shock curve (Fig. IV.12, (c)). Since ρ_ℓ is lower than ρ_r, the left state is also on the left side of the 2-Hugoniot curve issued from the right state: $\rho_\ell^\varepsilon < \rho_r^\varepsilon = h_+^\varepsilon(\theta_r^\varepsilon) < h_+^\varepsilon(\theta_\ell^\varepsilon)$. We again consider the increasing function $w_-^\varepsilon - w_+^\varepsilon$ on its domain of definition $[(i_-^\varepsilon)^{-1}(0), (i_+^\varepsilon)^{-1}(0)]$. This function is negative at θ_ℓ and positive at θ_r. So the intersection point of the two wave curves is the intersection of the two shock curves and $\widetilde{\rho} > \rho_\ell, \rho_r$. This corresponds to the first subcase of the third case of the theorem.

If ρ_ℓ is greater than ρ_r, the decreasing 2-shock curve issued from the right state divides the domain $[\rho_r, \rho^*] \times [0, \theta_r]$ and the same arguments as before lead to the result.

Case $\theta_\ell = \theta_r$ (Fig. IV.12, (d)). Assume that $\rho_r < \rho_\ell$. We again consider the increasing function $w_-^\varepsilon - w_+^\varepsilon$. It is positive at $\theta = \theta_\ell$ and negative for $\theta = (i_-^\varepsilon)^{-1}(0) < \theta_\ell$ (since $h_+^\varepsilon\left((i_-^\varepsilon)^{-1}(0)\right) > \rho_r > 0$). So it equals zero for a value $\theta < \theta_r$. So $\left(\widetilde{\rho}, \widetilde{\theta}\right)$ is the intersection of the 1-rarefaction curve and the 2-shock curve, which leads to the solution given in the first case of the theorem. The case $\rho_r > \rho_\ell$ is similar. ∎

F.2 Proof of proposition 20

Proof. 1. Consider the domain where $\{\rho = \rho_\ell\}$.

If the left state is connected to the intermediate state via a rarefaction wave, then this rarefaction fan is contained between the speeds $\lambda_-^\varepsilon = \cos\theta_\ell - \sqrt{\varepsilon p'(\rho_\ell)\rho_\ell}|\sin\theta_\ell|$ and $\tilde{\lambda}_-^\varepsilon = \cos\tilde{\theta} - \sqrt{\varepsilon p'(\tilde{\rho})\tilde{\rho}}|\sin\tilde{\theta}|$. Since $\lambda_-^\varepsilon < \cos\theta_\ell$, the domain $\{\rho = \rho_\ell\}$ cannot contain the contact wave with speed $\cos\theta_\ell$.

If the left state is connected to the intermediate state via a shock wave, then we have $\cos\tilde{\theta} < \cos\theta_\ell$ and $\rho_\ell < \tilde{\rho}$, which yields

$$s^- = \cos\tilde{\theta} + \rho_\ell \frac{(\cos\tilde{\theta} - \cos\theta_\ell)}{\tilde{\rho} - \rho_\ell} \quad < \cos\tilde{\theta}.$$

So the domain $\{\rho = \rho_\ell\}$ cannot contain the contact wave with speed $\cos\theta_\ell$.

So in both cases, the domain $\{\rho = \rho_\ell\}$ cannot contain the contact wave with speed $\cos\theta_\ell$. The same arguments show that the domain $\{\rho = \rho_r\}$ cannot contain the contact wave with speed $\cos\theta_r$.

Finally, we easily check that a contact wave with propagation speed $\cos\tilde{\theta}$ can occur within the intermediate domain $\{\rho = \tilde{\rho}\}$. Indeed, if the intermediate state is connected to the left state (resp. to the right state) via a rarefaction wave, then we have $\tilde{\lambda}_-^\varepsilon < \cos\tilde{\theta} < \tilde{\lambda}_+^\varepsilon$ and if the intermediate state is connected to the left state (resp. to the right state) via a shock wave then $s^- < \cos\tilde{\theta}$ (resp. $s^+ > \cos\tilde{\theta}$).

2. Like in the previous point, the contact wave can be located only in the intermediate state $\{\rho = 0\}$. But the propagation speed is not unique: it can be all the intermediate speeds between the two fans of rarefaction. ∎

G Proofs of lemma 22 and propositions 21, 23 and 24 (limits of solutions of the Riemann problem)

We recall that the quantities indexed by - (resp. by +) are implicitly those related to the left state (resp. the right state). The characteristic speeds related to the intermediate state will be denoted by $\widetilde{\lambda}^{\varepsilon}_{\pm}$.

G.1 Proof of proposition 21

Proof. (a) Let us suppose that $\rho_\ell < \rho_r$ (the opposite case is similar). According to theorem 19 and proposition 17, the intermediate state angle $\widetilde{\theta}^{\varepsilon}$ tends to θ_ℓ (since $\widetilde{\rho}^{\varepsilon}$ belongs to the interval $]\rho_\ell, \rho_r[$ for each ε). In addition, it is easy to check that the two speeds $\widetilde{\lambda}^{\varepsilon}_{+}$ and λ_+ tend to $\cos\theta_r$ as ε tends to 0. So the rarefaction wave turns into a contact wave with speed $\cos\theta_r$. As regards the shock wave, its speed is given by

$$s^{\varepsilon} = \frac{\widetilde{\rho}^{\varepsilon}\cos\widetilde{\theta}^{\varepsilon} - \rho_\ell\cos\theta_\ell}{\widetilde{\rho}^{\varepsilon} - \rho_\ell} = \cos\theta_\ell + \widetilde{\rho}^{\varepsilon}\frac{\cos\widetilde{\theta}^{\varepsilon} - \cos\theta_\ell}{\widetilde{\rho}^{\varepsilon} - \rho_\ell}, \tag{IV.63}$$

and tends to $\cos\theta_\ell$ as ε tends 0. Finally, the two waves coincide and make a single contact wave.

(b) As in the first case, it is easy to check that the two rarefaction waves (cf. theorem 19) turn into contact waves with speeds respectively equal to $\cos\theta_r$ and $\cos\theta_\ell$.

(c) According to proposition 16, $h^{\varepsilon}_{-}(\rho^{\varepsilon}_r)$ tends to $\theta_\ell < \theta_r$. So, theorem 19 implies that we are looking for the limit of the intersection point of the two shock curves issued from the left and the right states. These intermediate states $(\widetilde{\rho}^{\varepsilon}, \widetilde{\theta}^{\varepsilon})$ are the solutions of the non-linear system

$$H_\varepsilon(\rho_\ell, \theta_\ell, \widetilde{\rho}^{\varepsilon}, \widetilde{\theta}^{\varepsilon}) = [\Phi(\cos\theta) + \varepsilon p(\rho)]_\ell\,[\rho]_\ell - [\Psi(\cos\theta)]_\ell\,[\rho\cos\theta]_\ell = 0, \tag{IV.64}$$

$$H_\varepsilon(\widetilde{\rho}^{\varepsilon}, \widetilde{\theta}^{\varepsilon}, \rho_r, \theta_r) = [\Phi(\cos\theta) + \varepsilon p(\rho)]_r\,[\rho]_r - [\Psi(\cos\theta)]_r\,[\rho\cos\theta]_r = 0, \tag{IV.65}$$

$$\widetilde{\rho}^{\varepsilon} \geq \max(\rho_\ell, \rho_r),\ \widetilde{\theta}^{\varepsilon} \in]\theta_\ell, \theta_r[\,. \tag{IV.66}$$

From proposition 16, for all $\theta \in]\theta_\ell, \theta_r[$, the largest zero of the function $\rho \to H_\varepsilon(\rho_\ell, \theta_\ell, \rho, \theta)$ tends to ρ^* as $\varepsilon \to 0$. Indeed, if $\widetilde{\rho}^{\varepsilon}$ does not tend to ρ^*, $\widetilde{\theta}^{\varepsilon}$ simultaneously tend to θ_ℓ and to θ_r (which is different from θ_ℓ), which is absurd. Therefore, $\widetilde{\rho}^{\varepsilon}$ increases and tends to ρ^*. Besides, from equation (IV.64), $\varepsilon p(\widetilde{\rho}^{\varepsilon})$ is bounded as ε tends to 0 (because $\widetilde{\theta}^{\varepsilon}$ is bounded too) and we deduce that $\varepsilon p(\widetilde{\rho}^{\varepsilon})$ converges to a non-zero value $\bar{p}$. Finally, we can easily check that the system given in proposition 21 is equivalent to the limit of system (IV.64)-(IV.65)-(IV.66). ∎

G.2 Proof of lemma 22

Proof. Suppose that $\widetilde{\lambda}_+ = \lim \widetilde{\lambda}^{\varepsilon}_{+}$ is finite. Since $\varepsilon p(\rho^{\varepsilon}_r) \to \bar{p}_r > 0$, then $\varepsilon p'(\rho^{\varepsilon}_r) \to +\infty$ and consequently $\lambda_+ \to +\infty$. The limit rarefaction wave has a fan for speeds s belonging to $]\widetilde{\lambda}_+, +\infty[$. The 2-rarefaction wave satisfies, for all $s \in]\widetilde{\lambda}_+, +\infty[$,

$$s = \lambda_+(\rho(s), \theta(s)) = \cos(\theta(s)) + \sqrt{\varepsilon p'(\rho(s))\rho(s)}\,|\sin(\theta(s))|.$$

So, for a fixed s, $\rho(s) \to \rho^*$ as $\varepsilon \to 0$ and we have $(\rho^* - \rho(s)) = O(\varepsilon^{\frac{1}{\gamma+1}})$. Thus, $\varepsilon p(\rho(s))$ $(= O(\varepsilon^{1-\frac{\gamma}{\gamma+1}})) \to 0$ as $\varepsilon \to 0$. So the rarefaction wave tends to the combination of a shock wave between the states $(\widetilde{\rho}, \theta_r, 0)$ and $(\rho^*, \theta_r, 0)$ with speed $\widetilde{\lambda}_+$ and a declustering wave.

If $\widetilde{\rho} < \rho^*$, then $\widetilde{\lambda}_+$ equals 0 and the previous arguments apply. Let us look at the case $\widetilde{\rho} = \rho^*$. If $\widetilde{\lambda}_+$ is finite, then in the previous conclusion the shock wave disappears since the two states on both sides of the shock wave are equal. And it confirms that $\bar{\bar{p}} = \lim \varepsilon p(\widetilde{\rho}^\varepsilon)$ equals zero. If $\widetilde{\lambda}_+$ is infinite like $\lim \lambda_+^\varepsilon$, then the rarefaction wave turns into a shock wave with an infinite speed between the states $(\rho^*, \theta_r, \bar{p}_\ell)$ and $(\rho^*, \theta_r, \bar{p}_r)$. ∎

G.3 Proof of proposition 23

Proof. (a) The intermediate state is here the intersection of the 2-rarefaction curve and the 1-shock curve. The intersection state exists for all ε by the monotony of the two curves (cf. propositions 16 and 17). Let us note that there is no reason to have a finite limit of $\widetilde{\lambda}^\varepsilon$ since the intermediate state can tend to ρ^*. So let us consider several cases:

Case (i) $\widetilde{\rho^\varepsilon} \to \rho_\ell$. In this case, the 1-shock disapears and the solution is given by lemma 22.

Case (ii) $\widetilde{\rho^\varepsilon} \to \widetilde{\rho} \in]\rho_\ell, \rho^*[$. In this case, it is easy to check that the 1-shock becomes a 1-contact and the limit of the 2-rarefaction is given by lemma 22.

Case (iii) $\widetilde{\rho^\varepsilon} \to \rho^*$. Now let us look at the limit of $\varepsilon p(\widetilde{\rho}^\varepsilon)$. $(\widetilde{\rho^\varepsilon}, \widetilde{\theta^\varepsilon})$ satisfies (IV.64). The terms $[\rho \cos\theta]$ and $[\rho]$ are bounded and $\cos\tilde{\theta}$ tends to $\cos\theta_r$ (cf. third point of proposition 17), so taking the limit $\varepsilon \to 0$ in (IV.64), we get

$$[\varepsilon p(\rho)]_\ell \, [\rho]_\ell \to 0.$$

Thus $\varepsilon p(\widetilde{\rho}^\varepsilon)$ tends to 0 (since $[\rho]_\ell$ does not tend to 0). Finally lemma 22 applies and we conclude that the 2-rarefaction tends to a declustering wave. The shock speed, given by (IV.63), tends to $\cos\theta_\ell = \cos\theta_r$. So the 1-shock tends to a contact discontinuity.

(b) The limit of the 2-rarefaction is given by lemma 22 (with $\widetilde{\rho}^\varepsilon = 0$ and $\lambda_+ = \cos\theta_r$).The 1-rarefaction wave turns into a contact wave as before.

(c) We first consider the case where the intermediate state is the intersection point of two shock curves (for all ε, $h_-^\varepsilon(\theta_r^\varepsilon) > \rho_r^\varepsilon$): by the monotony of these curves, $\widetilde{\rho}^\varepsilon$ is larger than ρ_r^ε and so tends to ρ^* too. Besides, the intermediate state $(\widetilde{\rho}^\varepsilon, \widetilde{\theta}^\varepsilon)$ satisfies the system (IV.64)-(IV.66), which implies that $\varepsilon p(\widetilde{\rho}^\varepsilon)$ converges to a value denoted $\bar{\bar{p}}$. By taking the limit $\varepsilon \to 0$ in (IV.65), we obtain

$$[\Psi(\cos\theta)]_r \, [\cos\theta]_r = 0$$

(since $\tilde{\theta}$ is bounded and $[\rho]_r$ tends to 0), which implies that $\tilde{\theta}$ equals θ_r. Finally, if the limit $[\varepsilon p(\rho)]_r$ is non zero, the propagation speed is infinite (easily deduced from (IV.65)) and if it is zero, there is no discontinuity.

Consider now the case where the intermediate state is the intersection of the 1-shock curve issued from the left state and the 2-rarefaction curve issued from the right state (for all ε, $h_-^\varepsilon(\theta_r^\varepsilon) < \rho_r^\varepsilon$). From proposition 17 (point 3), the intermediate angle $\widetilde{\theta}^\varepsilon$ tends to θ_r. Besides, thanks to proposition 16, the intermediate density $\widetilde{\rho}^\varepsilon$ tends to ρ^*. From (IV.33), $\varepsilon p(\widetilde{\rho}^\varepsilon)$ converges to a value denoted by $\bar{\bar{p}}$ which is given by the limit Rankine

Hugoniot relation. So the limit intermediate state is $(\rho^*, \theta_r, \bar{\bar{p}})$. Finally lemma 22 applies: the rarefaction turns into a shock. ∎

G.4 Proof of proposition 24

Proof. (a) Since the intermediate density $\tilde{\rho}^\varepsilon$ is comprised between the left and right ones: $\rho_\ell^\varepsilon < \tilde{\rho}^\varepsilon < \rho_r^\varepsilon$ (cf. theorem 19) and since ρ_ℓ^ε, $\rho_r^\varepsilon \to \rho^*$, we also have $\tilde{\rho}^\varepsilon \to \rho^*$. For the intermediate angle $\tilde{\theta}^\varepsilon$, the previous proof shows that it tends to θ_r. Let us note that the 2-rarefaction wave tends to a contact wave (because $\lambda_{r+}^\varepsilon, \tilde{\lambda}_+^\varepsilon \to +\infty$). Let s be the limit of the 1-shock speed and let us note that s is lower than $\cos\theta_\ell$. Like in prop. 23, subcase (c), the intermediate pressure is equal to

$$\bar{\bar{p}} = \bar{p}_\ell + \lim \frac{[\rho\cos\theta]_\ell\,[\Psi]_\ell}{[\rho]_\ell}.$$

Now let us look at the limit of the shock speed. For finite ε, the shock speed is given by (IV.63). Since $\varepsilon p(\rho_r^\varepsilon) \to \bar{p}_r > 0$, we have $\varepsilon^{\frac{1}{\gamma}} = O(\rho^* - \rho_r^\varepsilon)$ and then $\varepsilon^{\frac{1}{\gamma}} = O(\rho^* - \tilde{\rho}^\varepsilon)$. On the other hand, from lemma 17 we have $(i^\varepsilon)^{-1}(\rho) - \theta_r = O(\varepsilon^{\frac{1}{2\gamma}})$ and therefore we get

$$\cos\tilde{\theta}^\varepsilon - \cos\theta_\ell = -2\sin\left(\frac{\tilde{\theta}^\varepsilon + \theta_\ell}{2}\right)\sin\left(\frac{\tilde{\theta}^\varepsilon - \theta_\ell}{2}\right) = O(\varepsilon^{\frac{1}{2\gamma}}).$$

Thus, we easily get that s^ε tends to $\cos\theta_\ell$ and then that the pressure $\bar{\bar{p}}$ equals $\bar{p}_\ell$.

(b) Here the proof is similar to the case where only one state converges to the congested state (see proof of prop. 23).

(c) Consider the case where the solution is the limit of two shock waves. By the monotony of the shock curves, the intermediate density is larger than the right and left ones (cf. theorem 19) and so it tends to the congested density too.

Suppose that the intermediate angle is not equal to θ_ℓ. The limit 1-shock speed, given by (IV.63), is $-\infty$. Besides, equation (IV.64) implies that $\varepsilon p(\tilde{\rho}^\varepsilon)$ tends to $+\infty$. Then $[\varepsilon p(\rho)]_r$ tends to $+\infty$ and from (IV.65), we deduce that the 2-shock speed has to tend to $+\infty$ too. If the intermediate angle tends to θ_ℓ, then it does not tend to θ_r and the same arguments apply. From (IV.64)-(IV.65), the quantities $[\varepsilon p(\rho)]_\ell[\rho]_\ell$ and $[\varepsilon p(\rho)]_r[\rho]_r$ are bounded and the limit of their quotient is

$$\frac{[\varepsilon p(\rho)]_r[\rho]_r}{[\varepsilon p(\rho)]_\ell[\rho]_\ell} \underset{\varepsilon\to 0}{\longrightarrow} \frac{[\Psi(\cos\theta)]_r[\cos\theta]_r}{[\Psi(\cos\theta)]_\ell[\cos\theta]_\ell}.$$

Besides, it is easily checked that

$$\frac{[\varepsilon p(\rho)]_r[\rho]_r}{[\varepsilon p(\rho)]_\ell[\rho]_\ell} \underset{\varepsilon\to 0}{\sim} \frac{\tilde{\rho} - \rho_r}{\tilde{\rho} - \rho_\ell} \underset{\varepsilon\to 0}{\sim} \frac{\rho^* - \rho_r}{\rho^* - \rho_\ell},$$

where the last equivalence results from the fact that $(\rho^* - \tilde{\rho}) = o(\varepsilon^{\frac{1}{\gamma}})$ since $\varepsilon p(\tilde{\rho}) \to +\infty$ and $(\rho^* - \rho_{\ell,r}) = O(\varepsilon^{\frac{1}{\gamma}})$, $\varepsilon^{\frac{1}{\gamma}} = O(\rho^* - \rho_{\ell,r})$. Finally, we have

$$\frac{\rho^* - \rho_r}{\rho^* - \rho_\ell} = \left(\frac{\varepsilon p(\rho_\ell)}{\varepsilon p(\rho_r)}\right)^{\frac{1}{\gamma}} \to \left(\frac{\bar{p}_\ell}{\bar{p}_r}\right)^{\frac{1}{\gamma}}.$$

Consider now the limit of a solution consisting of one shock wave and one rarefaction wave. From lemma 22, the intermediate angle $\widetilde{\theta}^{\varepsilon}$ tends to θ_r and from equation (IV.64), $[\varepsilon p(\widetilde{\rho}]_\ell\,[\rho]_\ell$ tends to a non zero value: $\rho^*\,[\Psi(\cos\theta)]_\ell\,[\cos\theta]_\ell$. Because $[\rho]_\ell$ tends to 0, $\varepsilon p(\widetilde{\rho}^{\varepsilon})$ has to tend to $+\infty$, which is absurd since $\varepsilon p(\widetilde{\rho}^{\varepsilon}) < \varepsilon p(\rho_r^{\varepsilon})$. ∎

H Smooth and incompressible vector field with values in $\mathbb{S}^1$

Proposition 29. *1. For any smooth vector field $u(x)$ on $\mathbb{R}^d$, the following idendity hold:*

$$\nabla_x \cdot ((u\cdot\nabla_x)u) = Tr(\nabla_x u(\nabla_x u)^T) + (u\cdot\nabla_x)(\nabla\cdot u),$$

where Tr is the trace of a matrix and T denotes the transpose operator.

2. Besides, if $u(x)$ takes value in $\mathbb{S}^1$, then we have $Tr(\nabla_x u(\nabla_x u)^T) = (\nabla_x\cdot u)^2$.

3. Therefore, if $u(x)$ is an incompressible smooth vector field with values in $\mathbb{S}^1$, then u satisfies:

$$\nabla_x \cdot ((u\cdot\nabla_x)u) = 0.$$

The last identity, combined with proposition 12, implies that an incompressible smooth vector field in $\mathbb{S}^1$ is a constant field.

Proof. 1- The first identity results from the following calculus :

$$\nabla_x \cdot ((u\cdot\nabla_x)u) = \sum_{i,j=1,..,d} \partial_i(u_j\partial_j u_i) = \sum_{i,j=1,..,d} \partial_i u_j \partial_j u_i + \sum_{i,j=1,..,d} u_j\partial_i\partial_j u_i.$$

2- In dimension 2, we have

$$Tr(\nabla_x u(\nabla_x u)^T) = (\partial_1 u_1)^2 + 2\partial_1 u_2\partial_2 u_1 + (\partial_2 u_2)^2.$$

Besides since $|u| = 1$, then we have

$$u_j\partial_1 u_j = 0, \quad u_j\partial_2 u_j = 0,$$

which imply

$$\partial_1 u_2 = -\frac{u_1\partial_1 u_1}{u_2}, \quad \partial_2 u_1 = -\frac{u_2\partial_2 u_2}{u_1}.$$

Then, we obtain:

$$Tr(\nabla_x u(\nabla u)^T) = (\partial_1 u_1)^2 + 2\partial_1 u_1\partial_2 u_2 + (\partial_2 u_2)^2 = (\nabla_x\cdot u)^2.$$

∎

I Numerical simulations of the Individual-Based model

In this appendix, we present preliminar numerical simulations of the Individual-Based-Model (IV.36)-(IV.37). In these numerical simulations, the forces ξ^a and ξ^r are normalized :

$$\xi_k^a = \frac{B_k^a}{|B_k^a|}, \quad \xi_k^r = \frac{B_k^r}{|B_k^r|},$$

where B_k^a and B_k^r are the barycenter of the mass distribution inside discs centered in position X_k and with radii R_a and R_r:

$$B_k^a = \frac{\sum_{j,|X_j - X_k| \leq R_a} X_j - X_k}{\sum_{j,|X_j - X_k| \leq R_a} 1}, \quad B_k^r = \frac{\sum_{j,|X_j - X_k| \leq R_r} X_j - X_k}{\sum_{j,|X_j - X_k| \leq R_r} 1}.$$

When B_k^a (resp. B_k^r) vanishes, the force ξ_k^a (resp. ξ_k^r) is not defined and so is supposed to be zero. Note that in appendix A, ξ_k^a and ξ_k^r was defined as the barycenter B_k^a and B_k^r without any renormalization. This change does not affect the macroscopic behaviour and enables to clearly separate the role of R_a and ν_a in the parameters analysis. The attraction frequency $\nu_k^a = \nu_a$ is chosen to be constant, while the repulsion frequency encodes the density constraint. Indeed, it is depending on a local evaluation of the volume fraction α_k^{loc}:

$$\nu_k^r = \nu_r \left(\alpha_k^{\text{loc}}\right),$$

where ν_r is an increasing function such that $\nu_r(\alpha) \to +\infty$ as α tends to 1. That means that the larger the density is, the faster the particle adjusts its direction. In the following study, we will consider the function :

$$\nu_r(\alpha) = \varepsilon \alpha p'(\alpha), \quad p(\alpha) = \left(\frac{1}{\alpha} - 1\right)^{-\gamma},$$

where $\gamma > 1$ and $\varepsilon \ll 1$, but this choice has no influence on the qualitative behaviour of the dynamics. The local volume fraction α_k^{loc} is given by:

$$\alpha_k^{\text{loc}} = \frac{\pi \delta^2 \sum_{j,|X_j - X_k| \leq R_r} 1}{\pi R_r^2},$$

where δ denotes the radius of the particles.

Such an attraction-repulsion model could lead to clustering, that is formation of groups. As clustering is concerned, the main parameters of this model are the two attraction parameters (the attractive frequency ν_a and the attractive radius R_a) and the two repulsive parameters (the repulsive radius R_r and the radius of the particles δ). Future studies will aim at determining the range of these parameters leading to clustering.

Here, we only illustrate what kind of patterns we obtain. We consider the movement of 1000 particles in a square box of size 1 with periodic boundary conditions. Their speed is one and the time step is $\Delta t = 10^{-3}$. Initially, the particles are uniformly distributed. The parameters are the following :

– the volume fraction α:

$$\alpha = \frac{\text{volume of the particles}}{\text{volume of the domain}} = \frac{N\pi\delta^2}{L^2},$$

which satisfies the upper-bound:

$$\alpha \leq 1.$$

It is equivalent to define the repulsive radius δ or the volume fraction α.

– The attraction frequency ν_a.

The repulsion radius R_r, which has to be greater than δ, will satisfy here : $R_r^2 = 10\delta^2$ (that means that about 10 neighbour particles can enter a repulsion disc). The attraction radius R_a is taken to be equal to 0.2, the parameters of the respulsive intensity ν_r are : $\gamma = 2$ and $\varepsilon = 10^{-4}$. In Fig. IV.13, we plot several results of simulations. At fixed α, we can see that the number of clusters is increasing with respect to ν_a, while their volume is decreasing. At fixed ν_a, the number of clusters increases and then decreases with respect to α. The investigation of the conditions for the formation of clusters with all the parameters will be carried out in a future work.

Bibliography

[1] I. Aoki, A simulation study on the schooling mechanism in fish, *Bulletin of the Japanese Society of Scientific Fisheries (Japan)*, 48 :1081-1088, 1982.

[2] D. Armbruster, P. Degond, C. Ringhofer, Continuum models for interacting machines, Networks of Interacting Machines: Production Organization in Complex Industrial Systems And Biological Cells, 1-32. D. Armbruster, K. Kaneko and A. Mikhailov (eds), World Sci., 2005.

[3] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, Proc. Nat. Acad. Sci. USA, 105 :1232, 2008.

[4] A. Barrat, E. Trizac, M.H. Ernst, Granular gases: dynamics and collective effects, *J. Phys. Condens. Matter*, 17 :S2429, 2005.

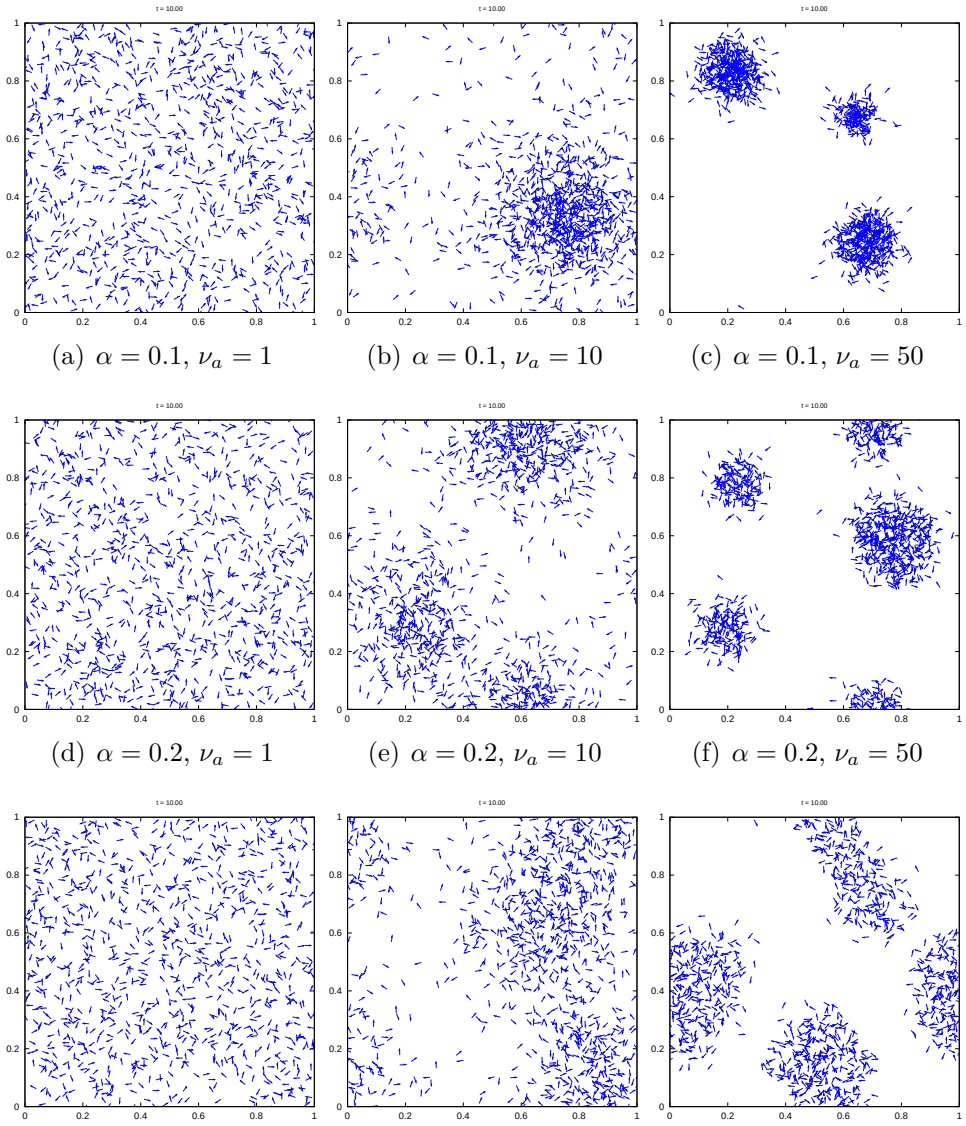
[5] N. Bellomo, C. Dogbe, On the modelling crowd dynamics from scaling to hyperbolic macroscopic models, *Math. Models Methods Appl. Sci.*, 18 :1317-1345, 2008.

[6] F. Berthelin, P. Degond, M. Delitala, M. Rascle, A model for the formation and the evolution of traffic jams, *Arch. Rational Mech. Anal.*, 187 :185-220, 2008.

[7] F. Berthelin, P. Degond, V. Le Blanc, S. Moutari, M. Rascle, J. Royer, A traffic-flow model with constraints for the modeling of traffic jams, *Math. Models Methods Appl. Sci.*, 18 :1269-1298, 2008.

[8] E. Bertin, M. Droz, G. Grégoire, Boltzmann and hydrodynamic description for self-propelled particles, *Phys. Rev. E.*, 74 :022101, 2006.

[9] F. Bouchut, On zero pressure gas dynamics, Advances in kinetic theory and computing: selected papers, *Ser. Adv. Math. Appl. Sci.*, 22 :171-190, 1994.



(a) $\alpha = 0.1$, $\nu_a = 1$ (b) $\alpha = 0.1$, $\nu_a = 10$ (c) $\alpha = 0.1$, $\nu_a = 50$

(d) $\alpha = 0.2$, $\nu_a = 1$ (e) $\alpha = 0.2$, $\nu_a = 10$ (f) $\alpha = 0.2$, $\nu_a = 50$

(g) $\alpha = 0.5$, $\nu_a = 1$ (h) $\alpha = 0.5$, $\nu_a = 10$ (i) $\alpha = 0.5$, $\nu_a = 50$

Figure IV.13: Numerical simulations of the Individual-Based model with 1000 particles at time $t = 10$. Other parameters : $R_a = 0.2$, $R_r^2 = 10\delta^2$, $c = 1$, $\Delta t = 10^{-3}$, $\varepsilon = 10^{-4}$, $\gamma = 2$.

[10] F. Bouchut, Y. Brenier, J. Cortes, J.-F. Ripoll, A hierarchy of models for two-phase flows, *J. Nonlinear Sci.*, 10 :639-660, 2000.

[11] J.A. Carrillo, M.R. D'Orsogna, V. Panferov, Double milling in self-propelled swarms from kinetic theory, *Kinet. Relat. Models*, 2 :363-378, 2009.

[12] J.A. Carrillo, M. Fornasier, J. Rosado, G. Toscani, Asymptotic flocking dynamics for the kinetic Cucker-Smale model, preprint

[13] C. Cercignani, R. Illner, M. Pulvirenti, *The mathematical theory of dilute gases*. Springer-Verlag, New-York, 1994.

[14] Y. Chuang, M.R. D'Orsogna, D. Marthaler, A.L. Bertozzi, L.S. Chayes, State transitions and the continuum limit for a 2D interacting self-propelled particle system, *Physica D*, 232 :33-47 (2007)

[15] I.D.Couzin, J. Krause, Self-organization and collective behavior in vertebrates, *Adv. Study Behav.*, 32 :1-75, 2003.

[16] I.D. Couzin, J. Krause, R. James, G.D. Ruxton, N.R. Franks, Collective memory and spatial sorting in animal groups, *J. Theoret. Biol.*, 218 :1-11, 2002.

[17] F. Cucker, S. Smale, Emergent behavior in flocks, *IEEE Trans. Automat. Control*, 52 :852-862, 2007.

[18] F. Cucker, S. Smale, On the mathematics of emergence, *Japan J. Math.*, 2 :197-227, 2007.

[19] M.R. D'Orsogna, Y.L. Chuang, A.L. Bertozzi, L.S. Chayes, Self-propelled particles with soft-core interactions: patterns, stability, and collapse, *Phys. Rev. Lett.*, 96 :104302, 2006.

[20] P. Degond, Macroscopic limits of the Boltzmann equation: a review, Modeling and computational methods for kinetic equations, 3-57. P. Degond, L. Pareschi, G. Russo (eds), Modeling and Simulation in Science, Engineering and Technology Series, Birkhäuser, 2003.

[21] P. Degond, M. Delitala, Modelling and simulation of vehicular traffic jam formation, *Kinet. Relat. Models*, 1 :279-293, 2008.

[22] P. Degond, S. Motsch, Continuum limit of self-driven particles with orientation interaction, *Math. Models Methods Appl. Sci.*, 18 :1193-1215, 2008.

[23] P. Degond, S. Motsch, Large scale dynamics of the Persistent Turning Angle model for fish behaviour, *J. Stat. Phys.*, 131 :989-1021, 2008.

[24] P. Degond, S. Motsch, Macroscopic limit of self-driven particles with orientation interaction, *C. R. Math. Acad. Sci. Paris*, 345 :555-560, 2007.

[25] J. Gautrais, C. Jost, M. Soria, A. Campo, S. Motsch, R. Fournier, S. Blanco, G. Theraulaz, Analyzing fish movement as a persistent turning walker, *J. Math. Biol.*, 58 :429-445, 2008.

[26] G. Grégoire, H. Chaté, Onset of collective and cohesive motion, *Phys. Rev. Lett.*, 92 :025702, 2004.

[27] S.-Y. Ha, E. Tadmor, From particle to kinetic and hydrodynamic descriptions of flocking, *Kinet. Relat. Models*, 1 :415-435, 2008.

[28] D. Helbing, Traffic and related self-driven many-particle systems, *Rev. Modern Phys.*, 73 :1067-1141, 2001.

[29] D. Helbing, P. Molnar, Social force model for pedestrian dynamics, *Math. Comput. Simul. Phys. Rev. E*, 51 :4282-4286, 1995.

[30] P.G. LeFloch, A.E. Tzavaras, Representation of Weak Limits and Definition of Nonconservative Products, *SIAM J. Math. Anal.*, 30 :1309-1342, 1999.

[31] R.J. LeVeque, *Numerical Methods for Conservation Laws*. Birkhäuser-Verlag, Basel, 1990.

[32] B. Maury, J. Venel, Handling of contacts in crowd motion simulations, Trafic and Granular Flow '07, 171-180. Springer, Berlin, 2009.

[33] B. Maury, J. Venel, A mathematical framework for a crowd motion model, *C. R. Acad. Sci. Paris, Ser I*, 346 :1245-1250, 2008.

[34] A. Mogilner, L. Edelstein-Keshet, A non-local model for a swarm, *J. Math. Biol.*, 38 :534-570, 1999.

[35] A. Mogilner, L. Edelstein-Keshet, L. Bent, A. Spiros, Mutual interactions, potentials and individual distance in a social aggregation, *J. Math. Biol.*, 47 :353-389, 2003.

[36] M. Moussaïd, S. Garnier, G. Theraulaz, D. Helbing, Collective information processing and pattern formation in swarms, flocks and crowds, *Topics in Cognitive Science*, 1 :469-497, 2009.

[37] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, G. Theraulaz, Experimental study of the behavioural mechanisms underlying self-organization in human crowds, *Proc. R. Soc. Ser. B*, 276 :2755-2762, 2009.

[38] F. Peruani, A. Deutsch, M. Bär, Nonequilibrium clustering of self-propelled rods, *Phys. Rev. E*, 74 :30904, 2006.

[39] M.H. Pillot, Etude et modélisation des déplacements collectifs spontanés chez le mouton mérinos d'Arles (ovis aries), Master thesis, 2006.

[40] M.H. Pillot, J. Gautrais, J. Gouello, P. Michelena, A. Sibbald, R. Bon, Moving together, incidental leaders and naïve followers, in preparation.

[41] T. Poschel, H.J. Herrmann, Size segregation and convection, *Europhys. Lett.*, 29 :123-128, 1995.

[42] C.W. Reynolds, Flocks, herds and schools: A distributed behavioral model, *Comp. Graph.*, 21 :25-34, 1987.

[43] D. Serre, *Systems of Conservation Laws I*. Cambridge Univ. Press, Cambridge, 1999.

[44] D. Slepcev, Coarsening in nonlocal interfacial systems, *SIAM J. Math. Anal.*, 40 :1029-1048, 2008.

[45] H. Spohn, *Large Scale Dynamics of Interacting Particles*. Springer-Verlag, Berlin, 1991.

[46] C.M. Topaz, A.L. Bertozzi, M.A. Lewis, A Nonlocal Continuum Model for Biological Aggregation, *Bull. Math. Biol.*, 68 :1601-1623, 2006.

[47] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, O. Shochet, Novel Type of Phase Transition in a System of Self-Driven Particles, *Phys. Rev. Lett.*, 75 :1226-1229, 1995.

Chapitre V

Simulations numériques du système d'Euler avec congestion

Ce chapitre a été écrit en collaboration avec Pierre Degond et Jiale Hua. La suite de ce chapitre est écrite en anglais.

Numerical simulations of the Euler system with congestion constraint

1 Introduction

Several models involve congestion constraints: concentration constraints occur in multi-phase flow modeling [7], maximal density constraints occur when dealing with finite-size interactive agents in herds of gregarious mamals [11], in cars or pedestrians flows [2, 4, 5], flux constraints occur for supply chains,... Congested regions exhibit quite different dynamics than free regions. To study the transitions between congested and free regions, a general methodology was carried out in [4, 11]: the stiffness of the constraint leads to a singular pertubative problem and then the limit problem provides a clear cut-off between the two dynamics. In this paper, we will consider the Euler system with a singular pressure which encodes a maximal density constraint. As in [11], the limit problem is a two-phase model between incompressible regions, where the maximal density is reached, and compressible regions for lower densities. Our goal is to provide numerical schemes that are able to capture this limit problem and these transitions. In this paper, we adapt and compare two numerical methods presented in [12] and in [10].

A lot of effort have been made to devise numerical schemes valid for all Mach numbers, that is for both compressible and incompressible flows. They avoid the switch between different methods, when different Mach numbers occur in different sub-domain. Most methods are based on a time semi-implicit discretization of the fluxes (the mass flux and the pressure part of the momentum flux are taken implicitly) associated with the resolution of the elliptic equation satisfied by the pressure. The implicit treatment of the pressure flux enables stability with respect to the propagation of fast acoustic waves in the low-Mach number limit but induces a lot of diffusion. Such schemes have been developped in the framework of essentially two approaches. The first approach is the extension of incompressible methods to compressible flows and leads to pressure correction methods

on staggered grids [14] and their conservative versions (SIMPLE method [22]). Further developments are the multiple pressure variables methods [20]. On the other hand, the second approach is the extension of compressible conservative methods to incompressible flows and leads to preconditioning techniques [13] and flux splitting methods with the introduction of implicitness [17]. The methods we consider in this paper are among the simplest one: the scheme in [12] is a semi-implicit scheme with a division of the pressure into explicit and an implicit parts and in [10], the Gauge decomposition of the momentum enables the hydrostatic pressure to act only on the divergence-free part of the momentum. The former method will be called in the present paper the Direct method, while the Gauge method will refer to the latter.

The Direct and the Gauge methods are called asymptotic-preserving (AP) since they are uniformly consistant with the limit problem. Besides, they are also uniformly stable. These methods are expected to capture both the compressible and the incompressible dynamics arising in the congestion limit of the Euler system with the maximal density constraint. In this case, such AP numerical schemes are very powerful since they provide the dynamics of transitions, for which analytical results may sometimes be lacking. Moreover, they enable to avoid to deal with physical and numerical interface tools, such as front-tracking [23] or volume-of-fluid methods [16]. Unlike these tracking methods, ours are front-capturing methods and then share some analogies with level-set [21] and diffusive interface methods [1]: like level-set method, the dynamics of the transition are implicitly embodied in the dynamics of an auxiliary function which is here the density and like the diffusive interface methods, the sharp interface is viewed as the limit of the smooth transitions of the perturbative problem.

The paper is organized as follows. In section 2, we introduce the Euler system with the maximal density constraint, we give its formal asymptotic limit and we describe one-dimensional solutions. In section 3, we describe the time semi-discretization of the Direct and Gauge schemes. The fully space and time discretizations are exposed in section 4: they are based on the centered Rusanov sheme, also called the local Lax-Friedrichs scheme. Finally, numerical simulations are performed in section 5 to compare the two schemes.

2 The Euler system with congestion and its asymptotic limit

2.1 The model

We consider the two-dimensional Euler system:

$$\partial_t \rho + \nabla_x \cdot q = 0, \tag{V.1}$$

$$\partial_t q + \nabla_x \cdot \left(\frac{q \otimes q}{\rho} \right) + \nabla_x p(\rho) = 0, \tag{V.2}$$

where $\rho(x,t) \in \mathbb{R}$ denotes the mass density, $q = \rho u(x,t) \in \mathbb{R}^2$ is the momentum density depending on the position $x \in \mathbb{R}^2$ and the time $t > 0$. The pressure $p(\rho)$ is an increasing function such that $p(\rho) \sim \rho^\gamma$ for densities $\rho \ll 1$ and $p(\rho) \to +\infty$ as ρ tends to the

congestion density ρ^*. In the following, we will consider the function:

$$p(\rho) = \frac{1}{\left(\frac{1}{\rho} - \frac{1}{\rho^*}\right)^\gamma}. \tag{V.3}$$

This pressure prevents the density from exceeding the congestion density: it is a covolume equation of state. A variant is the van der Waals equation of state [15]. The operators ∇_x and $\nabla_x \cdot$ are the gradient and the divergence of vector fields or tensor. For two vectors a and b, $a \otimes b$ denotes the tensor product.

This model already appears in [11] with the additional constraint: $q/\rho = u \in \mathbb{S}^1$. In this paper, we focus on the pressure singularity and the adapted numerical schemes.

The singular pressure induces two different dynamics: for regions with densities near ρ^*, the pressure takes very large values in comparison with the pressure in low-density regions. As in [11] and previous work on traffic modeling [4], we would like to clearly separate the two different dynamics. To this aim, we rescale $p(\rho)$ into $\varepsilon p(\rho)$, where the parameter $\varepsilon \ll 1$ is the scale of the pressure in the low density regions: $p(\rho) = O(\varepsilon)$ for density $\rho \ll 1$ while $p(\rho) = O(1)$ for density $\rho \sim \rho^*$. Denoting ρ^ε and q^ε the new unknowns, system (V.1)-(V.2) becomes:

$$\partial_t \rho^\varepsilon + \nabla_x \cdot q^\varepsilon = 0, \tag{V.4}$$

$$\partial_t q^\varepsilon + \nabla_x \cdot \left(\frac{q^\varepsilon \otimes q^\varepsilon}{\rho^\varepsilon}\right) + \nabla_x \varepsilon p(\rho^\varepsilon) = 0. \tag{V.5}$$

This is a strictly hyperbolic problem for positive density, with caracteristic speeds in the x_1-direction (where x_1 is the first component of a basis of $\mathbb{R}^2$) given by:

$$\lambda_1^\varepsilon = u_{x_1}^\varepsilon - \sqrt{\varepsilon p'(\rho^\varepsilon)}, \quad \lambda_2^\varepsilon = u_{x_1}^\varepsilon, \quad \lambda_3^\varepsilon = u_{x_1}^\varepsilon + \sqrt{\varepsilon p'(\rho^\varepsilon)}, \tag{V.6}$$

where $u_{x_1}^\varepsilon$ is the x_1-component of the macroscopic velocity $u^\varepsilon(x,t) = q^\varepsilon(x,t)/\rho^\varepsilon(x,t)$ is the macroscopic velocity. Standard hyperbolic numerical schemes have to resolve the CFL condition:

$$\max(|\lambda_1^\varepsilon|, |\lambda_2^\varepsilon|, |\lambda_3^\varepsilon|)\Delta t \leq \Delta x. \tag{V.7}$$

We will see in the next section that this constraint may be too more stringent for these schemes to capture the asymptotic limit.

2.2 The asymptotic limit

The limit of the pressure term $\varepsilon p(\rho^\varepsilon)$ depends on the limit of ρ^ε. Indeed, if $\rho^\varepsilon \to \rho$ with $\rho < \rho^*$, then $\varepsilon p(\rho^\varepsilon)$ converges to 0. Otherwise, $\rho^\varepsilon \to \rho^*$ and the limit of $\varepsilon p(\rho^\varepsilon)$, denoted $\bar{p}$, can be non zero and depends on the convergence speed of ρ^ε. We assume that the limit $\bar{p}$ is always finite. Therefore, the formal limit of system (V.4)-(V.5) as ε goes to zero is:

$$\partial_t \rho + \nabla_x \cdot q = 0, \tag{V.8}$$

$$\partial_t q + \nabla_x \cdot \left(\frac{q \otimes q}{\rho}\right) + \nabla_x \bar{p} = 0, \tag{V.9}$$

$$(\rho - \rho^*)\bar{p} = 0. \tag{V.10}$$

A one-dimensional version of this asymptotic model was proposed for two-phase flow modeling in [7], where the density plays the role of the volume fraction of liquid in a liquid-gaz model. The derivation of the model lies on a relaxation to zero of the relative velocities of the gaz and liquid and is therefore different from the one studied in this article. Existence and stability of solutions are proved for the one-dimensional version of system (V.8)-(V.9)-(V.10) in [3], and in all dimensions with viscous term in [19].

As regards the caracteristic speeds, we note that if $\varepsilon p(\rho^\varepsilon)$ tends to $\bar{p} < +\infty$, then we have $\rho^* - \rho^\varepsilon = O(\varepsilon^{\frac{1}{\gamma}})$ and then $\varepsilon p'(\rho^\varepsilon) = O(\varepsilon^{\frac{1}{\gamma}-1})$. Therefore, if $\gamma > 1$, λ_1^ε and λ_3^ε can become infinite and waves with infinite speed can occur. It is the low-Mach number asymptotics leading to incompressible dynamics. Actually, in the congested domain where $\rho = \rho^*$, system (V.8)-(V.9)-(V.10) yields the incompressible Euler equation:

$$\begin{aligned} &\rho = \rho^*, \\ &\nabla_x \cdot u = 0, \\ &\partial_t u + u \cdot \nabla_x u + \frac{1}{\rho^*} \nabla_x \bar{p} = 0. \end{aligned} \tag{V.11}$$

Equation (V.11) is the incompressible constraint and the Lagrange multiplier related to this constraint is the pressure $\bar{p}$. The CFL condition (V.7) degenerates into $\Delta t = 0$: standard hyperbolic schemes are unable to compute the asymptotic dynamics. Thus, numerical schemes with relaxed CLF condition have to be designed.

In the free domain where $\rho < \rho^*$, the CFL condition (V.7) is not an obstacle although the system degenerates into a non-hyperbolic problem: $\lim \lambda_1^\varepsilon = \lim \lambda_3^\varepsilon = u_{x_1}$. This is a large-Mach number asymptotic. Numerical schemes, originally developped for hyperbolic systems, have to be proved to capture this singular limit. System (V.8)-(V.9)-(V.10) yields the pressureless gaz dynamics:

$$\begin{aligned} &\rho < \rho^*, \\ &\partial_t \rho + \nabla_x \cdot q = 0, \\ &\partial_t q + \nabla_x \cdot \left(\frac{q \otimes q}{\rho} \right) = 0. \end{aligned}$$

Without upper-bound on the density, this system would lead to concentration phenomena even for smooth initial data and so the density may become a measure with singular part. It is related to the so-called sticky particle dynamics. Existence of solutions and numerical schemes have been developped in the one-dimensional case [6, 8].

To capture this asymptotic limit, we need to devise numerical schemes valid for low and large Mach number. One-dimensional Riemann problems will be test-cases to valid them. Therefore, in the next section, we briefly give the theoretical solutions of the Riemann problem related to the asymptotic problem (V.8)-(V.9)-(V.10).

2.3 Solutions of the one-dimensional problem: the Riemann problem and the cluster collisions

The one-dimensional version of system (V.4)-(V.5) can be written as follows:

$$\partial_t \rho + \partial_x q = 0, \tag{V.12}$$

$$\partial_t q + \partial_x \left(\frac{q^2}{\rho} \right) + \varepsilon \partial_x p(\rho) = 0, \tag{V.13}$$

where $q(x,t)$ is now a scalar function and x is the position in $\mathbb{R}$. In this section, we first investigate the Riemann problem and the limits of its solutions as $\varepsilon \to 0$ and then we briefly recall the one-dimensional solutions of the limit system (V.8)-(V.9)-(V.10) provided in [7], which consists of the collision of two finite clusters (domain where $\rho = \rho^*$).

2.3.1 The one-dimensional Riemann problem

The Riemann problem is an initial value problem, where the initial condition is a discontinuity between two constant states:

$$(\rho_0, q_0)(x) = \begin{cases} (\rho_\ell, q_\ell), & \text{for } x < 0, \\ (\rho_r, q_r), & \text{for } x \geq 0. \end{cases}$$

The solutions of this problem for the Euler system (V.12)-(V.13) are well known. So, the strategy is to take the limits of these solutions as ε goes to zero. This provides the solutions of the Riemann problem for the singular asymptotic limit (V.8)-(V.9)-(V.10).

Riemann problem for system (V.12)-(V.13). The two caracteristic speeds $\lambda_\pm^\varepsilon$ and the two caracteristic fields $r_\pm^\varepsilon$ of the one-dimensional system (V.12)-(V.13) are:

$$\lambda_\pm^\varepsilon = u \pm \sqrt{\varepsilon p'(\rho)}, \quad r_\pm^\varepsilon = \begin{pmatrix} 1 \\ u \pm \sqrt{\varepsilon p'(\rho)} \end{pmatrix}.$$

It can be easily checked that both characteristic fields are genuinely non linear for positive densities. Therefore, the solutions of the Riemann problem are made of constant states separated by rarefaction or shock waves [18]. The integral curves $i_\pm^\varepsilon$ issued from a given state $(\hat{\rho}, \hat{q})$ are:

$$i_\pm^\varepsilon(\rho) = \rho \hat{u} \pm \rho \sqrt{\varepsilon} (P(\rho) - P(\hat{\rho})), \tag{V.14}$$

where P is an antiderivative function of $\sqrt{p'(u)}/u$ and $\hat{u} = \hat{q}/\hat{\rho}$. The Hugoniot curves $h_\pm^\varepsilon$ issued from $(\hat{\rho}, \hat{q})$ are derived from the Rankine-Hugoniot conditions and lead to:

$$h_\pm^\varepsilon(\rho) = \hat{u} \pm \sqrt{\frac{\rho}{\hat{\rho}}} \sqrt{(\rho - \hat{\rho})(\varepsilon p(\rho) - \varepsilon p(\hat{\rho}))}. \tag{V.15}$$

The admissible shock speeds between states $(\hat{\rho}, \hat{q})$ and $(\rho, q_\pm^s(\rho))$ are given by:

$$\sigma_\pm = \frac{(q_\pm^s(\rho) - \hat{q})}{(\rho - \hat{\rho})} = \hat{u} \pm \sqrt{\frac{\rho}{\hat{\rho}}} \sqrt{\frac{(\varepsilon p(\rho) - \varepsilon p(\hat{\rho}))}{(\rho - \hat{\rho})}}. \tag{V.16}$$

Geometric considerations on the intersection of the integral and Hugoniot curves enable us to solve the Riemann problem [18]. These arguments are really simplified in the limit $\varepsilon \to 0$, due to the much simpler behaviour of the integral and Hugoniot curves: they both converge to the union of the straight lines $\{q = \rho\hat{u}\}$ and $\{\rho = \rho^*\}$ (see appendix A.1).

The limits of the solutions to the Riemann problem. Let us state the main results. We defer all the proofs to appendix A. We introduce the following initial conditions:

$$(\rho_0^\varepsilon, q_0^\varepsilon)(x) = \begin{cases} (\rho_\ell^\varepsilon, q_\ell^\varepsilon), & \text{for } x < 0, \\ (\rho_r^\varepsilon, q_r^\varepsilon), & \text{for } x \geq 0, \end{cases}$$

with $(\rho_\ell^\varepsilon, q_\ell^\varepsilon, \varepsilon p(\rho_\ell^\varepsilon)) \to (\rho_\ell, q_\ell, \bar{p}_\ell)$ and $(\rho_r^\varepsilon, q_r^\varepsilon, \varepsilon p(\rho_r^\varepsilon)) \to (\rho_r, q_r, \bar{p}_r)$ as ε goes to zero. The following proposition provides the solution when $\rho_\ell, \rho_r < \rho^*$ and so $\bar{p}_\ell = \bar{p}_r = 0$. The nature of the solution depends on the sign of the relative velocity $u_\ell - u_r$, where $u_\ell = q_\ell/\rho_\ell$ and $u_r = q_r/\rho_r$:

Proposition 30. ***(case** $\rho_\ell < \rho^*, \rho_r < \rho^*$**)***

1. *If $u_\ell - u_r < 0$, then the solution consists of two contact waves connecting the two states to the vacuum. This is summerized in the following diagram:*

$$(\rho_\ell, q_\ell, 0) \overset{contact}{\longrightarrow} (0, q_\ell, 0) \overset{vacuum}{\longrightarrow} (0, q_r, 0) \overset{contact}{\longrightarrow} (\rho_r, q_r, 0).$$

2. *If $u_\ell - u_r > 0$, then the solution consists of two shock waves connecting the left state $(\rho_\ell, q_\ell, 0)$ to an intermediate state $(\rho^*, \tilde{q}, \bar{p})$ and then $(\rho^*, \tilde{q}, \bar{p})$ to the right state $(\rho_r, q_r, 0)$:*

$$(\rho_\ell, q_\ell, 0) \overset{shock}{\longrightarrow} (\rho^*, \tilde{q}, \bar{p}) \overset{shock}{\longrightarrow} (\rho_r, q_r, 0),$$

where the intermediate momentum $\tilde{q}$ and the intermediate pressure $\bar{p}$ are:

$$\tilde{q} = u_\ell \rho^* - \sqrt{\frac{\rho^*}{\rho_\ell}}\sqrt{(\rho^* - \rho_\ell)\bar{p}} = u_r \rho^* + \sqrt{\frac{\rho^*}{\rho_r}}\sqrt{(\rho^* - \rho_r)\bar{p}},$$

$$\bar{p} = (u_\ell - u_r)^2 \left(\sqrt{\frac{\rho^* - \rho_\ell}{\rho_\ell \rho^*}} + \sqrt{\frac{\rho^* - \rho_r}{\rho_r \rho^*}} \right)^{-2},$$

and the shock speeds σ_- and σ_+ are given by:

$$\sigma_- = u_\ell - \sqrt{\frac{\rho^*}{\rho_\ell(\rho^* - \rho_\ell)}}\sqrt{\bar{p}}, \quad \sigma_+ = u_r + \sqrt{\frac{\rho^*}{\rho_r(\rho^* - \rho_r)}}\sqrt{\bar{p}}.$$

3. *If $u_\ell = u_r$, then the solution consists of only one contact wave connecting $(\rho_\ell, q_\ell, 0)$ to $(\rho_r, q_r, 0)$:*

$$(\rho_\ell, q_\ell, 0) \overset{contact}{\longrightarrow} (\rho_r, q_r, 0).$$

The proof of this proposition is deferred in appendix A.2.1. It is the same results as those obtained in [7], where they directly proved it by defining a notion of entropy solutions for the asymptotic problem.

We now consider the case where the left state is a congested state: $\rho_\ell = \rho^*, \rho_r < \rho^*$ and $\bar{p}_\ell < +\infty$. By a symmetry argument, the case $\rho_\ell < \rho^*, \rho_r = \rho^*$ can be easily deduced. The limits of rarefaction waves when one state tends to congestion lead to the so-called declustering waves:

Definition 2. ***(Declustering waves)*** *A declustering wave consists in a shock wave between two congested state, with infinite speed, and with a zero pressure for positive time.*

The following proposition states the solutions of the Riemann problem.

Proposition 31. ***(case*** $\rho_\ell = \rho^*, \rho_r < \rho^*$***)***

1. *If* $u_\ell - u_r < 0$*, then the solution consists of one declustering wave connecting the left state* $(\rho^*, q_\ell, \bar{p}_\ell)$ *to a congested and pressureless state* $(\rho^*, q_\ell, 0)$ *and then a contact wave connecting* $(\rho^*, q_\ell, 0)$ *to vacuum and another contact wave connecting vacuum to the right state* $(\rho_r, q_r, 0)$*:*

$$(\rho^*, q_\ell, \bar{p}_\ell) \overset{declust.}{\longrightarrow} (\rho^*, q_\ell, 0) \overset{contact}{\longrightarrow} (0, q_\ell, 0) \overset{vacuum}{\longrightarrow} (0, q_r, 0) \overset{contact}{\longrightarrow} (\rho_r, q_r, 0).$$

2. *If* $u_\ell - u_r > 0$*, then the solution consists of two shock waves connecting the left state* $(\rho^*, q_\ell, \bar{p}_\ell)$ *to an intermediate congested state* $(\rho^*, \tilde{q}, \bar{p})$ *and then connecting this intermediate state to the right state* $(\rho_r, q_r, 0)$*:*

$$(\rho^*, q_\ell, \bar{p}_\ell) \overset{shock}{\longrightarrow} (\rho^*, \tilde{q}, \bar{p}) \overset{shock}{\longrightarrow} (\rho_r, q_r, 0),$$

where the intermediate momentum $\tilde{q}$ *and the intermediate pressure* $\bar{p}$ *are given by:*

$$\tilde{q} = \rho^* u_\ell, \quad \bar{p} = \frac{\rho^* \rho_r}{\rho^* - \rho_r}(u_\ell - u_r)^2,$$

and the two speeds σ_- *and* σ_+ *are:*

$$\sigma_- = -\infty, \quad \sigma_+ = u_r + \sqrt{\frac{\rho^*}{\rho_r}}\sqrt{\frac{\bar{p}}{\rho^* - \rho_r}}.$$

3. *If* $u_\ell = u_r$*, then the solution consists of one declustering wave connecting the left state* $(\rho^*, q_\ell, \bar{p}_\ell)$ *to the intermediate state* $(\rho^*, q_\ell, 0)$ *and one contact wave connecting the intermediate state to the right state* $(\rho_r, q_r, 0)$*:*

$$(\rho^*, q_\ell, \bar{p}_\ell) \overset{declust.}{\longrightarrow} (\rho^*, q_\ell, 0) \overset{contact}{\longrightarrow} (\rho_r, q_r, 0).$$

The proof of this proposition is given in appendix A.2.2.

Finally, we consider the case where both left and right asymptotic states are congested: $\rho_\ell = \rho_r = \rho^*$. Besides, we assume that $\bar{p}_\ell$ and $\bar{p}_r$ are finite.

Proposition 32. ***(case*** $\rho_\ell = \rho^*, \rho_r = \rho^*$***,*** $\rho_\ell^\varepsilon > \rho_r^\varepsilon$***)***

1. *If* $u_\ell - u_r < 0$*, then the solution consists of one declustering wave connecting the left state* $(\rho^*, q_\ell, \bar{p}_\ell)$ *to a congested and pressureless state* $(\rho^*, q_\ell, 0)$*, then two contact waves connecting* $(\rho^*, q_\ell, 0)$ *to vacuum and then vacuum to* $(\rho^*, q_r, 0)$ *and another declustering wave connecting* $(\rho^*, q_r, 0)$ *to the right state* $(\rho_r, q_r, 0)$*:*

$$(\rho^*, q_\ell, \bar{p}_\ell) \overset{declust.}{\longrightarrow} (\rho^*, q_\ell, 0) \overset{contact}{\longrightarrow} (0, q_\ell, 0) \overset{vacuum}{\longrightarrow} (0, q_r, 0) \overset{contact}{\longrightarrow} (\rho^*, q_r, 0) \overset{declust.}{\longrightarrow} (\rho^*, q_r, \bar{p}_r).$$

2. *If $u_\ell - u_r > 0$, then the solution consists of two shock waves with infinite propagation speed connecting the left state $(\rho^*, q_\ell, \bar{p}_\ell)$ to an intermediate congested state $(\rho^*, \tilde{q}, +\infty)$ with infinite pressure and then this intermediate state to the right state $(\rho^*, q_r, \bar{p}_r)$:*

$$(\rho^*, q_\ell, \bar{p}_\ell) \quad \overset{shock}{\longrightarrow} \quad (\rho^*, \tilde{q}, +\infty) \quad \overset{shock}{\longrightarrow} \quad (\rho^*, q_r, \bar{p}_r),$$

where $\tilde{q}$ is the unique solution of:

$$\frac{[q]_\ell}{[q]_r} = \left(\frac{\bar{p}_r}{\bar{p}_\ell}\right)^{\frac{1}{2\gamma}}.$$

3. *If $u_\ell = u_r$, then the solution consists of a uniform constant state $(\rho^*, q_r, \bar{p}_r)$.*

The proof of this proposition is in appendix A.2.3.

2.3.2 The one-dimensional cluster collisions

The Riemann problem where both the left and right states are congested is ill-posed since infinite pressure may appear to correct the discontinuity in the pressure in the incompressible domain. Like in [3, 7, 11], we have to restrict to the collision of finite congested domains. Consider two one-dimensional clusters which collide at a time t_c. Before collision, the left (resp. right) cluster at time $t < t_c$ extends between $a_\ell(t)$ and $b_\ell(t)$ (resp. $a_r(t)$ and $b_r(t)$) and moves with speed:

$$u_\ell = a'_\ell(t) = b'_\ell(t) \quad (\text{resp. } u_r = a'_r(t) = b'_r(t)).$$

After collision, the two clusters agregate and form a new cluster at time $t > t_c$ extending between $a(t)$ and $b(t)$ and moving with speed $u = a'(t) = b'(t)$. Therefore, ρ and u are given for $t < t_c$ by:

$$\rho = \rho^* 1_{[a_\ell(t), b_\ell(t)]} + \rho^* 1_{[a_r(t), b_r(t)]}, \quad u = u_\ell 1_{[a_\ell(t), b_\ell(t)]} + u_r 1_{[a_r(t), b_r(t)]},$$

and for $t > t_c$ by:

$$\rho = \rho^* 1_{[a(t), b(t)]}, \quad u = u 1_{[a(t), b(t)]},$$

where 1_I denotes the indicator function of interval I. We denote by $m = b_\ell(t_c) = a_r(t_c)$ the collision point. We look for a pressure written as $\bar{p}(x,t) = \pi(x)\delta(t - t_c)$. Conditions to have such kind of solution for the one-dimensional version of (V.8)-(V.9)-(V.10) was obtained in [7]. We report them in the following proposition.

Proposition 33. *1- Supposing that $\bar{p}(x,t) = \pi(x)\delta(t - t_c)$ where π is continuous and zero outside the clusters, then u and π have to satisfy*

$$(u - u_\ell)(m - a(t_c)) + (u - u_r)(b(t_c) - m) = 0, \tag{V.17}$$

$$\pi(x) = \begin{cases} \rho^*(u - u_\ell)(m - x) \\ \quad +\rho^*(u - u_r)(b(t_c) - m), & \text{if } x \in [a(t_c), m], \\ \rho^*(u - u_r)(b(t_c) - x), & \text{if } x \in [m, b(t_c)], \end{cases} \tag{V.18}$$

2 - Under conditions (V.17)-(V.18), $(\rho, \rho u, \bar{p})$ is a solution (in a distributional sense) to the one-dimensional version of system (V.8)-(V.9)-(V.10).

3 Time semi-discretization schemes

3.1 The time semi-implicit discretization

In this paper, we shall apply asymptotic-preserving (AP) schemes developped in [12] and [10] to simulate the model. To start with, we first define a time semi-implicit discretization, which will be the building block of the considered (AP) schemes.

Let ρ^n, q^n be the approximations of ρ, q at $t^n = n\Delta t, n = 0, 1, \ldots$, where Δt is the time step. The semi-discretization of the AP scheme for the n-th time step is as follows:

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^{n+1} = 0, \tag{V.19}$$

$$\frac{q^{n+1} - q^n}{\Delta t} + \nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} \right) + \varepsilon \nabla_x p(\rho^{n+1}) = 0. \tag{V.20}$$

The full discretization in time and space is postponed to the next section.

Observe that the explicit part of the above scheme is pressureless. However, the pressureless Euler model is weakly hyperbolic. Thus we split the pressure term into explicit and implicit parts.

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^{n+1} = 0, \tag{V.21}$$

$$\frac{q^{n+1} - q^n}{\Delta t} + \nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} \right) + \varepsilon \nabla_x p_0(\rho^n) + \varepsilon \nabla_x p_1(\rho^{n+1}) = 0, \tag{V.22}$$

where the explicit part is given by:

$$p_0(\rho) = \begin{cases} \frac{1}{2} p(\rho), & \rho \le \rho_* - \delta, \\ \frac{1}{2} \Big\{ p(\rho_* - \delta) + p'(\rho_* - \delta)(\rho - \rho_* + \delta) & \\ \qquad + \frac{1}{2} p''(\rho_* - \delta)(\rho - \rho_* + \delta)^2 \Big\}, & \rho > \rho_* - \delta, \end{cases} \tag{V.23}$$

and the implicit part is:

$$p_1(\rho) = p(\rho) - p_0(\rho), \ \delta = \varepsilon^{\frac{1}{\gamma+2}}. \tag{V.24}$$

The choice of δ makes sure that p_0 and its derivatives up to second order are always bounded. To make sure all the coefficients appearing in the elliptic equation we shall derive in the next section are continuous, we choose p_0 to be a second order approximation to p, instead of a first order one. For later usage, also note that the function $p_1(\rho)$ is invertible. This is easily seen from the property of functions p and p_0, see Fig V.1. By the definition, the Courant-Friedrich-Lewy (CFL) condition for the explicit part is

$$\Delta t \le \frac{\sigma \Delta x}{\max\{|\frac{q}{\rho}| + \sqrt{\varepsilon p_0'(\rho)}\}}, \tag{V.25}$$

where σ is the Courant number. Since $\varepsilon p_0'$ is always bounded, the CFL condition can be satisfied uniformly in ε.

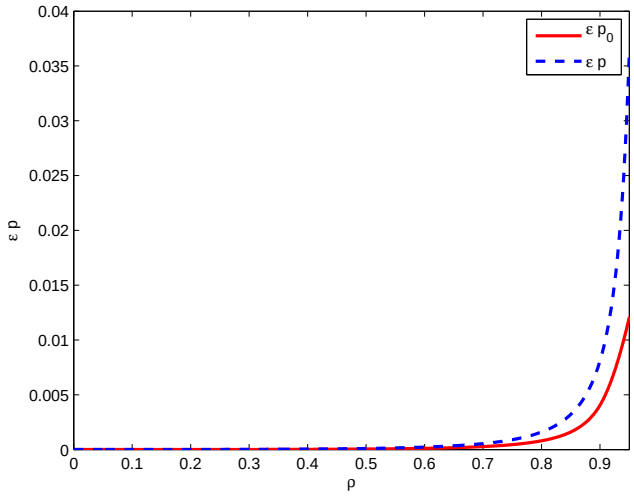


Figure V.1: The plots of εp and εp_0 as functions of ρ with $\varepsilon = 10^{-4}, \gamma = 2$.

3.2 The direct method

To get the solution, we shall rewrite the above scheme into another form. By inserting (V.22) into (V.21), we can get an elliptic equation for ρ:

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^n - \Delta t \varepsilon \Delta_x(p_1(\rho^{n+1})) - \Delta t \Delta_x \left(\frac{q^n \otimes q^n}{\rho^n} + \varepsilon p_0(\rho^n) \right) = 0. \tag{V.26}$$

From this equation, one can solve ρ^{n+1}. However, if we solve ρ directly, the density constraint $\rho \leq \rho^*$ may not be satisfied due to discretization errors. Thus, we write $\rho^{n+1} = \rho(p_1^{n+1})$ in (V.26) and solve this equation in terms of p_1. The density constraint $\rho \leq \rho^*$ will be automatically satisfied. Moreover, the positivity of ρ can be ensured by the fact that the discretized equation satisfies the maximum principle.

Once ρ^{n+1} is obtained, one can obtain q^{n+1} from (V.22) easily.

$$q^{n+1} = q^n - \Delta t \left[\nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} + \varepsilon p_0(\rho^n) \right) + \varepsilon \nabla_x (p_1(\rho^{n+1})) \right]. \tag{V.27}$$

We want to show that the system (V.26) and (V.22) becomes the incompressible Euler system as $\varepsilon \to 0$ in the congested region. However, in contrast with the low-Mach number limit of the isentropic Euler equation discussed in [12] or [10], the singularity in our model is embedded in the definition of p. New congestion regions may arise from non-congested ones. And by the analysis in section 2.3, there is the possibility that $\rho^\varepsilon \to \rho^*$ but the limit of the pressure $\varepsilon p_1(\rho^\varepsilon) \to 0$ as $\varepsilon \to 0$. This means that although the congestion density is reached, there is no real congestion in this region. So it seems better to characterize the congestion region by defining $\bar{p} = \lim_{\varepsilon \to 0} \varepsilon p_1(\rho^\varepsilon) > 0$. Currently, we can only show that in regions where both $\bar{p}^{n+1} = \lim_{\varepsilon \to 0} \varepsilon p_1((\rho^{n+1})^\varepsilon) > 0$ and $\bar{p}^n = \lim_{\varepsilon \to 0} \varepsilon p_1((\rho^n)^\varepsilon) > 0$, (V.26) and (V.27) tend to the incompressible Euler system as $\varepsilon \to 0$. The assumption $\bar{p}^{n+1} = \lim_{\varepsilon \to 0} \varepsilon p_1((\rho^{n+1})^\varepsilon) > 0$ is somehow essential given that declustering waves may appear in our model. The assumption $\bar{p}^n = \lim_{\varepsilon \to 0} \varepsilon p_1((\rho^n)^\varepsilon) > 0$ is also needed in case

of new congestion regions. In regions where $\bar{p}^{n+1} > 0$ and $\bar{p}^n > 0$, we have naturally that $(\rho^{n+1})^\varepsilon \to \rho^*$ and $(\rho^n)^\varepsilon \to \rho^*$ as $\varepsilon \to 0$. Taking the divergence of (V.27) and using (V.26), we can indeed recover (V.21), which leads to the incompressibility of $(q^{n+1})^\varepsilon$: $\nabla_x \cdot (q^{n+1})^\varepsilon = 0$. Here the implicitness in (V.21) is crucial. Then (V.11) follows. Although we can only prove the AP property inside a congestion region, the numerical solutions provide evidence that the scheme is globally AP, including at transition between compressible and incompressible region.

Remark. In the numerical simulation, we can also improve the accuracy by implementing a fully implicit scheme, which iterates the above scheme to solve (V.21) and (V.22) with $\frac{q\otimes q}{\rho}$ implicit. Suppose $\rho^{n+1,0} = \rho^n$ and $q^{n+1,0} = q^n$ and $\rho^{n+1,k}$, $q^{n+1,k}$ are the solutions to the following equations.

$$\frac{\rho^{n+1,k+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^n - \Delta t \varepsilon \Delta_x (p_1(\rho^{n+1,k+1})) - \Delta t \Delta_x \left(\frac{q^{n+1,k} \otimes q^{n+1,k}}{\rho^{n+1,k}} + \varepsilon p_0(\rho^n) \right) = 0, \tag{V.28}$$

$$q^{n+1,k+1} = q^n - \Delta t \left[\nabla_x \cdot \left(\frac{q^{n+1,k} \otimes q^{n+1,k}}{\rho^{n+1,k}} + \varepsilon p_0(\rho^n) \right) + \varepsilon \nabla_x (p_1(\rho^{n+1,k+1})) \right]. \tag{V.29}$$

As $k \to \infty$, the solution approximates to the one solving the fully implicit scheme (both in ρ and $\frac{q\otimes q}{\rho}$). This modification provides little improvement compared to the additional computational cost.

3.3 Gauge method

Another way to implement the AP scheme is the Gauge method developed in [10]. It can be obtained by applying the Gauge decomposition

$$q = a - \nabla_x \varphi, \quad \nabla_x \cdot a = 0, \tag{V.30}$$

where a is the incompressible part of field q and φ is the irrotationnal one. This decomposition is expected to be more robust for capturing incompressibility constraint. We will see that it is partially right. By including it into equations (V.21) and (V.22), we get:

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_x \cdot q^n - \Delta t \varepsilon \Delta_x (p_1(\rho^{n+1})) - \Delta t \Delta_x \left(\frac{q^n \otimes q^n}{\rho^n} + \varepsilon p_0(\rho^n) \right) = 0, \tag{V.31}$$

$$\Delta_x \varphi^{n+1} = \frac{1}{\Delta t}(\rho^{n+1} - \rho^n), \tag{V.32}$$

$$\Delta_x P^{n+1} = -\nabla_x^2 : \left(\frac{q^n \otimes q^n}{\rho^n} \right) - \varepsilon \Delta_x (p_0(\rho^n)), \tag{V.33}$$

$$\frac{a^{n+1} - a^n}{\Delta t} + \nabla_x \cdot \left(\frac{q^n \otimes q^n}{\rho^n} \right) + \varepsilon \nabla_x p_0(\rho^n) + \nabla_x P^{n+1} = 0, \tag{V.34}$$

$$q^{n+1} = a^{n+1} - \nabla_x \varphi^{n+1}, \tag{V.35}$$

where ∇_x^2 denotes the tensor of the second derivatives and for two tensors A and B, $A : B$ denotes the contracted product of tensor. Indeed, the equation (V.31) for p_1^{n+1} is derived

from (V.21) and (V.22) similarly as in the direct method. The Laplace equation (V.32) for φ^{n+1} is the direct consequence of applying the decomposition (V.35) and $\nabla_x \cdot a^{n+1} = 0$ to the density equation (V.21). The equation (V.33) for P^{n+1} and the equation (V.34) for a^{n+1} are obtained by inserting (V.35) and $\nabla_x \cdot a^{n+1} = \nabla_x \cdot a^n = 0$ into the momentum equation (V.22). Here a new unknown P is introduced, which is defined by

$$P^{n+1} = \varepsilon p_1(\rho^{n+1}) - \frac{\varphi^{n+1} - \varphi^n}{\Delta t}, \tag{V.36}$$

since (V.31)-(V.32) and (V.34) imply (V.22).

The original equations (V.21) and (V.22) can also be recovered from (V.31)-(V.35) by assuming $\nabla_x \cdot a^n = 0$. In fact, the equations for P^{n+1} and a^{n+1} ((V.33)-(V.34)) and $\nabla_x \cdot a^n = 0$ imply $\nabla_x \cdot a^{n+1} = 0$. This leads to the density equation (V.21) from the φ^{n+1} equation (V.32) and the decomposition (V.35). (V.21) combined with the p_1^{n+1} equation (V.31) will then allow us to recover the momentum equation (V.22).

The boundary conditions. By solving equations (V.31)-(V.35) in order, we can update the value of ρ, q. To do this, we need to provide boundary conditions for the Laplace equations for φ and P. The boundary conditions for P^{n+1} are somehow straightforward due to the implicit relation (V.36), once φ is known. And in solving φ^{n+1}, we may impose Dirichlet boundary condition on φ^{n+1} as follows:

$$\varphi^{n+1}|_\Omega = 0. \tag{V.37}$$

Indeed, other non-homogeneous Dirichlet boundary conditions can be chosen. This makes the unknown φ^{n+1} determined up to a linear function in space. However, this uncertainty can be removed by redefining a^{n+1}. So we can always impose the homogeneous Dirichlet boundary condition.

Simplification in the one dimensional case. Observe that in the 1D case, the incompressibility constraint $\nabla_x \cdot a = 0$ implies that a is constant in space. We may thus rewrite (V.33) and (V.34) into a simpler equation by using (V.36) :

$$a^{n+1} = a^n - \frac{\Delta t}{c-b}\left(\frac{(q^n)^2}{\rho^n} + \varepsilon p_0(\rho^n) + \varepsilon p_1(\rho^{n+1})\right)\Big|_b^c + \frac{1}{c-b}\left(\varphi^{n+1}|_b^c - \varphi^n|_b^c\right), \tag{V.38}$$

where the space-domain is the interval $[b, c]$ and for any function f, $f|_b^c = f(c) - f(b)$ denotes the difference between the left and right value of f at the boundary. This equation can be further simplified, since we impose the homogeneous Dirichlet boundary condition on φ all the time. This leads to:

$$a^{n+1} = a^n - \frac{\Delta t}{c-b}\left(\frac{(q^n)^2}{\rho^n} + \varepsilon p_0(\rho^n) + \varepsilon p_1(\rho^{n+1})\right)\Big|_b^c. \tag{V.39}$$

With this reformulation, in the 1D case, we can reduce the number of elliptic equations to be solved to one and update the space independent variable a more efficiently. Also as a consequence of (V.37), a^n should be defined as the average of q^n.

$$a^n = \frac{1}{c-b}\int_b^c q^n dx. \tag{V.40}$$

In summary, the Gauge method in the 1D case is implemented through (V.31), (V.32), (V.35), (V.39). And the scheme is still AP since it is a direct consequence of (V.21) and (V.22).

4 Full time and space discretization in 1D

The direct method. In the following, we consider the domain $[b, c] = [0, 1]$. Let the uniform spatial mesh be $\Delta x = \frac{1}{M}$, where M is a positive integer. Denote by $U_j^n = (\rho_j^n, q_j^n)$ the vector of the approximations of ρ, q at time $t^n = n\Delta t$ and positions $x_j = j\Delta x, j = 0, 1, \ldots, M$. We fully discretize the scheme (V.21) and (V.22) in the spirit of a local Lax-Friedrichs method as follows:

$$\begin{aligned}
&\frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} + \frac{1}{\Delta x}\Big[Q_{j+1/2}(U_j^n, U_{j+1}^n, U_j^{n+1}, U_{j+1}^{n+1}) \\
&\qquad\qquad - Q_{j-1/2}(U_{j-1}^n, U_j^n, U_{j-1}^{n+1}, U_j^{n+1})\Big] = 0, \\
&\frac{q_j^{n+1} - q_j^n}{\Delta t} + \frac{1}{\Delta x}\Big[F_{j+1/2}(U_j^n, U_{j+1}^n) - F_{j-1/2}(U_{j-1}^n, U_j^n)\Big] \\
&\qquad + \frac{1}{2\Delta x}\Big[\varepsilon p_1(\rho_{j+1}^{n+1}) - \varepsilon p_1(\rho_{j-1}^{n+1})\Big] = 0.
\end{aligned}$$

where the fluxes are given by:

$$\begin{aligned}
Q_{j+1/2}^{n+1/2} &= \frac{1}{2}\Big[q_j^{n+1} + q_{j+1}^{n+1}\Big] - \frac{1}{2}C_{j+1/2}^n(\rho_{j+1}^n - \rho_j^n), \\
F_{j+1/2}^n &= \frac{1}{2}\left[\frac{(q_j^n)^2}{\rho_j^n} + \frac{(q_{j+1}^n)^2}{\rho_{j+1}^n} + \varepsilon p_0(\rho_{j+1}^n) + \varepsilon p_0(\rho_j^n)\right] \\
&- \frac{1}{2}C_{j+1/2}^n(q_{j+1}^n - q_j^n), \\
C_{j+1/2}^n &= \max\left\{\left|\frac{q_j^n}{\rho_j^n}\right| + \sqrt{\varepsilon p_0'(\rho_j^n)}, \left|\frac{q_{j+1}^n}{\rho_{j+1}^n}\right| + \sqrt{\varepsilon p_0'(\rho_{j+1}^n)}\right\}.
\end{aligned}$$

The quantity $C_{j+1/2}^n$ is the local maximal characteristic speed. The quantity $C_{j+1/2}^n$ is the numerical viscosity which is needed for the scheme stability.

The Gauge method. Based on this discretization, one can apply the same strategy as described in section 3.2 to get an elliptic equation for ρ by substituting (4) into (4):

$$\begin{aligned}
&\frac{\rho_j^{n+1} - \rho_j^n}{\Delta t} + \frac{q_{j+1}^n - q_{j-1}^n}{2\Delta x} - \frac{\Delta t}{4\Delta x^2}\Big[\varepsilon p_1(\rho_{j+2}^{n+1}) - 2\varepsilon p_1(\rho_j^{n+1}) + \varepsilon p_1(\rho_{j-2}^{n+1})\Big] \\
&- \frac{1}{2\Delta x}\Big[C_{j+1/2}(\rho_{j+1}^n - \rho_j^n) - C_{j-1/2}(\rho_j^n - \rho_{j-1}^n)\Big] \\
&- \frac{\Delta t}{2\Delta x^2}\Big[F_{j+3/2}^n - F_{j+1/2}^n - F_{j-1/2}^n + F_{j-3/2}^n\Big] = 0.
\end{aligned} \tag{V.41}$$

This equation is consistent with eq.(V.26) of the direct method. Then we get a nonlinear equation

$$\begin{aligned}&\rho_j^{n+1} - \frac{\Delta t^2}{4\Delta x^2}\left[\varepsilon p_1(\rho_{j+2}^{n+1}) - 2\varepsilon p_1(\rho_j^{n+1}) + \varepsilon p_1(\rho_{j-2}^{n+1})\right]\\ =&\rho_j^n - \frac{\Delta t}{2\Delta x}(q_{j+1}^n - q_{j-1}^n) + \frac{\Delta t}{2\Delta x}\left[C_{j+1/2}(\rho_{j+1}^n - \rho_j^n) - C_{j-1/2}(\rho_j^n - \rho_{j-1}^n)\right]\\ &+ \frac{\Delta t^2}{2\Delta x^2}\left[F_{j+3/2}^n - F_{j+1/2}^n - F_{j-1/2}^n + F_{j-3/2}^n\right].\end{aligned} \tag{V.42}$$

As mentioned before, we shall use Newton iterations to solve this nonlinear equation and get p_1^{n+1}. The density ρ^{n+1} is then obtained by inverting the nonlinear function $p_1 = p_1(\rho)$ with another Newton iteration. Once ρ^{n+1} is solved, q^{n+1} can be obtained by

$$q_j^{n+1} = \Phi(U_{j-1}^n, U_j^n, U_{j+1}^n) - \frac{\Delta t}{2\Delta x}\left[\varepsilon p_1(\rho_{j+1}^{n+1}) - \varepsilon p_1(\rho_{j-1}^{n+1})\right], \tag{V.43}$$

with

$$\Phi(U_{j-1}^n, U_j^n, U_{j+1}^n) = q_j^n - \frac{\Delta t}{\Delta x}\left[F_{j+1/2}^n - F_{j-1/2}^n\right]. \tag{V.44}$$

Numerical diffusion. An important issue about the scheme is the numerical diffusion. From equation (V.42), it can be seen that the diffusion for ρ is of the order of

$$\left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p_0'(\rho^n)})\Delta x + \Delta t\varepsilon p_1'(\rho^n)\right)\Delta\rho^n + \Delta t\Delta(\rho^n u^n \otimes u^n). \tag{V.45}$$

And similarly, by inserting (4) into the pressure p_1 term in (4) one can see that the diffusion for q is of the order of

$$\left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p_0'(\rho^n)})\right)\Delta x\Delta q^n + \Delta t\varepsilon p_1'(\rho^n)\Delta q^n. \tag{V.46}$$

To damp out the oscillations in the mass and momentum equations, the requested numerical diffusion is

$$\left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p'(\rho^n)})\right)\Delta x\Delta\rho^n, \left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p'(\rho^n)})\right)\Delta x\Delta q^n, \tag{V.47}$$

respectively. To ensure that this numerical diffusion is achieved, we need the condition:

$$\left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p_0'(\rho^n)})\Delta x + \varepsilon p_1'(\rho^n)\Delta t\right) \geq \left(\frac{1}{2}(|u^n| + \sqrt{\varepsilon p'(\rho^n)})\right)\Delta x, \tag{V.48}$$

which leads to

$$\frac{\Delta t}{\Delta x} \geq \frac{1}{2\left(\sqrt{\varepsilon p_0'(\rho^n)} + \sqrt{\varepsilon p_1'(\rho^n)}\right)}. \tag{V.49}$$

This condition is automatically satisfied in the congested region ($\rho \to \rho^*$) for small ε, since $\varepsilon p_1'(\rho^n) \to \infty$ as $\varepsilon \to 0$. However, it contradicts the CFL condition in the non-congested region for small ε, since $\varepsilon p_1'(\rho^n)$ and $\varepsilon p_0'(\rho^n) \to 0$ as $\varepsilon \to 0$. From this analysis, there

should be no oscillations in the congestion region while the numerical diffusion may not be sufficient in the non-congested region. However, numerical simulations seem to indicate that the numerical viscosity in this scheme is sufficient to damp out the oscillations in the congested regions.

Also based on (4) and (4), we can have the full time-space discretization of the Gauge method. Indeed, for the Gauge method, we need to discretize (V.42). This leads to:

$$\begin{aligned}\frac{1}{\Delta x^2}\left[\varphi_{j+2}^{n+1}-2\varphi_j^{n+1}+\varphi_{j-2}^{n+1}\right]=&\frac{1}{\Delta t}(\rho_j^{n+1}-\rho_j^n)\\&-\frac{1}{2\Delta x}\left[C_{j+1/2}(\rho_{j+1}^n-\rho_j^n)-C_{j-1/2}(\rho_j^n-\rho_{j-1}^n)\right],\end{aligned}\tag{V.50}$$

$$\begin{aligned}a^{n+1}=&a^n-\Delta t\left(\frac{(q^n\otimes q^n)}{\rho^n}+\varepsilon p_0(\rho^n)+\varepsilon p_1(\rho^{n+1})\right)\Big|_0^1\\&+\frac{\Delta t}{2}\sum_1^M\left[C_{j+1/2}(q_{j+1}^n-q_j^n)-C_{j-1/2}(q_j^n-q_{j-1}^n)\right],\end{aligned}\tag{V.51}$$

$$q_j^{n+1}=a^{n+1}-\frac{1}{2\Delta x}(\varphi_{j+1}^{n+1}-\varphi_{j-1}^{n+1}).\tag{V.52}$$

The above equations are the direct consequences of (V.31), (V.32), (V.35), (V.39). However, in the numerical simulations, we will also test the schemes with (V.50) replaced by

$$\begin{aligned}\frac{1}{\Delta x^2}\left[\varphi_{j+1}^{n+1}-2\varphi_j^{n+1}+\varphi_{j-1}^{n+1}\right]=&\frac{1}{\Delta t}(\rho_j^{n+1}-\rho_j^n)\\&-\frac{1}{2\Delta x}\left[C_{j+1/2}(\rho_{j+1}^n-\rho_j^n)-C_{j-1/2}(\rho_j^n-\rho_{j-1}^n)\right],\end{aligned}\tag{V.53}$$

which may be justified as being the direct discretization of (V.32) with some numerical viscosity added to the right hand side. The stencils are different in two cases. We call the Gauge method with stencil (V.50) *Gauge 2* method and the one with stencil (V.53) *Gauge 1* method, since they use grids $\varphi_{j\pm2}$ and $\varphi_{j\pm1}$ respectively in addition to φ_j. There is a big difference between the two discretizations. In fact, we shall see in the next section that the Gauge 2 method (with (V.50)) yields almost the same numerical result as the direct method, while the Gauge 1 method (with (V.53)) performs quite differently from the direct method.

5 Numerical results

In this section, we use several numerical examples to test the performance of the schemes. Corresponding to different situations, four examples are tested. All these examples are the compositions of Riemann problems. Since the exact solutions to Riemann problems can be determined as in section 2, we can compare the exact and numerical solutions.

In the following, we choose $\gamma=2$ and the maximal density $\rho^*=1$. The test problems are

$$(P1) : (\rho, q)(x, 0) = \begin{cases} (0.7, 0.8), & x \in [0, 0.5), \\ (0.7, -0.8), & x \in (0.5, 1], \end{cases} \tag{V.54}$$

$$(P2) : (\rho, q)(x, 0) = \begin{cases} (0.7, 0.8), & x \in [0, 0.25), \\ (0.7, -0.8), & x \in (0.25, 0.75), \\ (0.7, 0.8), & x \in (0.75, 1], \end{cases} \tag{V.55}$$

$$(P3) : (\rho, q)(x, 0) = \begin{cases} (0.7, 0.8), & x \in [0, 0.25), \\ (0.8, -0.3), & x \in (0.25, 0.75), \\ (0.7, -1.2), & x \in (0.75, 1], \end{cases} \tag{V.56}$$

$$(P4) : (\rho, q)(x, 0) = \begin{cases} (0.8, 0.3), & x \in [0, 0.5), \\ (0.5, 0.1), & x \in (0.5, 1]. \end{cases} \tag{V.57}$$

The first example $(P1)$ illustrates how the AP schemes capture shocks near congestion. The second example $(P2)$ shows how the AP schemes work near vacuum. The third example $(P3)$ simulates the interaction of two shocks near congestion. The last example $(P4)$ shows some problems in the Gauge 1 method.

Example 1. The solution to the Riemann problem $(P1)$ consists of two shocks propagating in the opposite directions. The density of the intermediate state is close to the maximal density. In the following, we will test the two methods described in section 4 with different parameters ε and different mesh sizes $\Delta x, \Delta t$.

1. First, we choose $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. We shall compare the performances of the two methods proposed in section 4. It can be seen from Figure V.2 that there is large oscillations of the momentum in the congested region. But the propagation of the big shock is captured well. In comparison, the Gauge 1 method as illustrated in Figure V.2 eliminates all the oscillations.
2. We look how the choices of the parameters ε and Δx and Δt affect the numerical result. We may fix $\Delta x, \Delta t$ but change the value of ε so as to test the cases $\varepsilon < \Delta t$ and $\varepsilon > \Delta t$. From the numerical results, it can be seen that the oscillations in the momentum always appear for different choices of ε but are smaller as $\varepsilon \to 0$. This is just the AP property. As for the Gauge 1 method, it has the same performance for all values of ε. Thus it shares the same property.

Example 2. The solution to the problem $(P2)$ consists of two Riemann problems: one is the Riemann problem in (P1) and the other one is obtained by exchanging the left and right states. So in the second Riemann problem there are two rarefaction waves and a vacuum state appears as the intermediate state. As shown in proposition 30, these two rarefaction waves tend to be contact waves.

As in example 1, we choose $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. It can be seen from Figure V.4 that despite the large oscillations of the momentum in the congested region, the direct method captures the vacuum and rarefaction waves well. In comparison, the Gauge 1 method as illustrated in Figure V.4 show large diffusion.

Example 3. The solution to the problem $(P3)$ consists of two Riemann problems: both of them are like the Riemann problem in (P1). So there are two congested regions

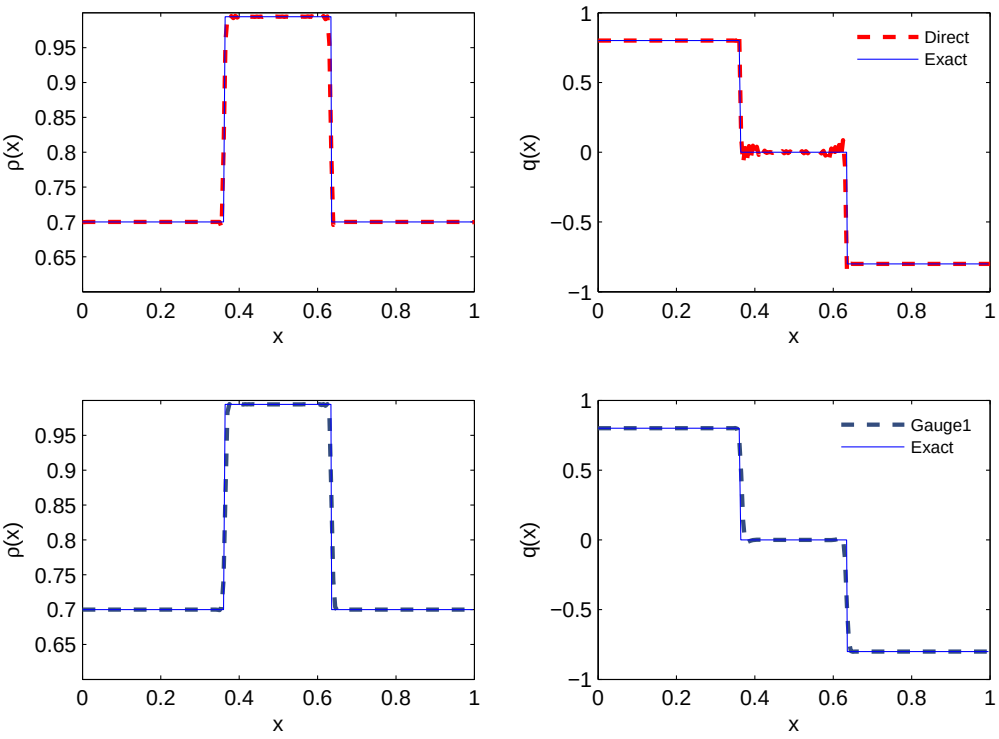
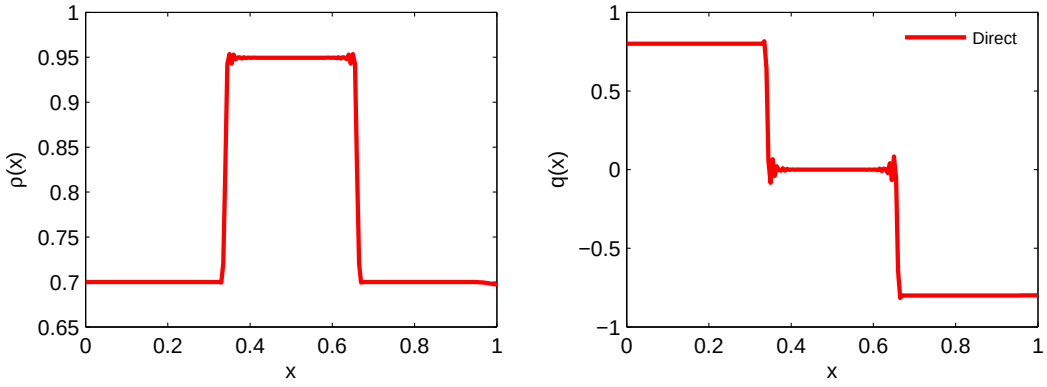
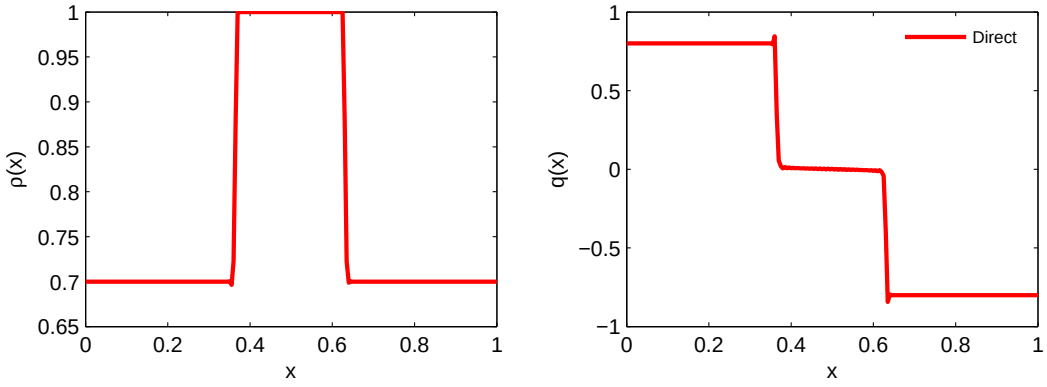


Figure V.2: The numerical results of the direct and Gauge 1 methods for problem (P1) at $t = 0.05$ with $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. The solid lines are the exact solution. The dashed curves are the numerical solutions. The left panels are for ρ, the right ones for q, both as a function of x.



(a) $\varepsilon = 10^{-2}$



(b) $\varepsilon = 10^{-8}$

Figure V.3: Fix $\Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. The numerical results for the direct scheme for problem (P1) at $t = 0.05$ with $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-8}$. The left panels are for ρ, the right ones for q, both as a function of x.

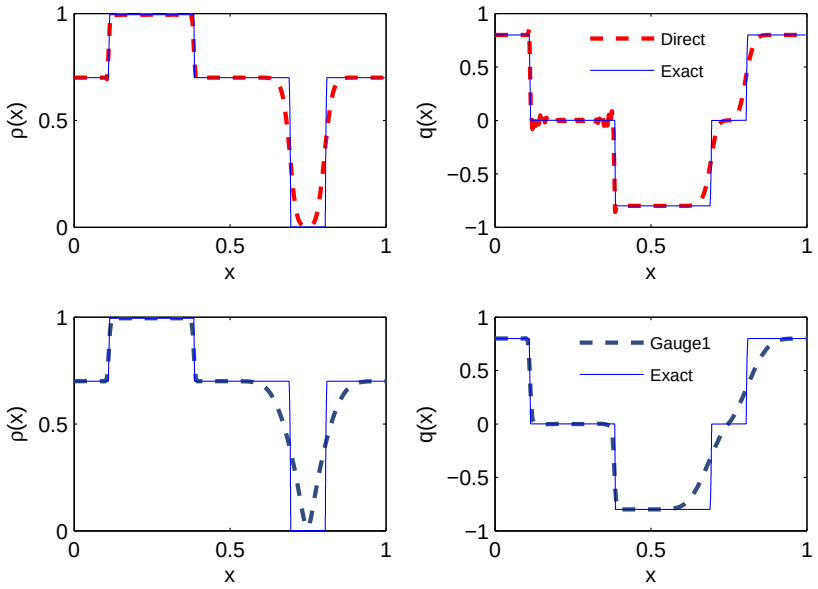


Figure V.4: The direct and the Gauge 1 methods for problem (P2) at $t = 0.05$ with $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. The solid lines are the exact solution. The dashed curves are the numerical solutions. The left panels are for ρ, the right ones for q, both as a function of x.

and eventually they will collide. We are interested in observing how the numerical methods behave at collision. We fix $\Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$ and choose $\varepsilon = 10^{-4}$ and 10^{-8}. Since only shocks are involved, we will use the Gauge 1 method only.

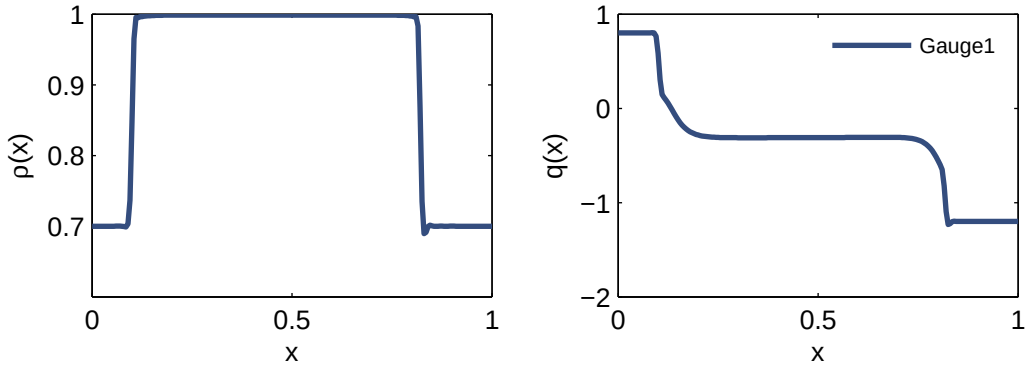
From Figures V.5 and V.6, it can be seen that as $\varepsilon \to 0$, the collision of these two congested shocks aggregate instantaneously.

Example 4. The solution to the Riemann problem $(P4)$ consists of two shocks with intermediate state away from the congestion density. So in the second Riemann problem there are two rarefaction waves and a vacuum state appears as the intermediate state. As shown in section 2, these two rarefaction waves tend to be contact waves.

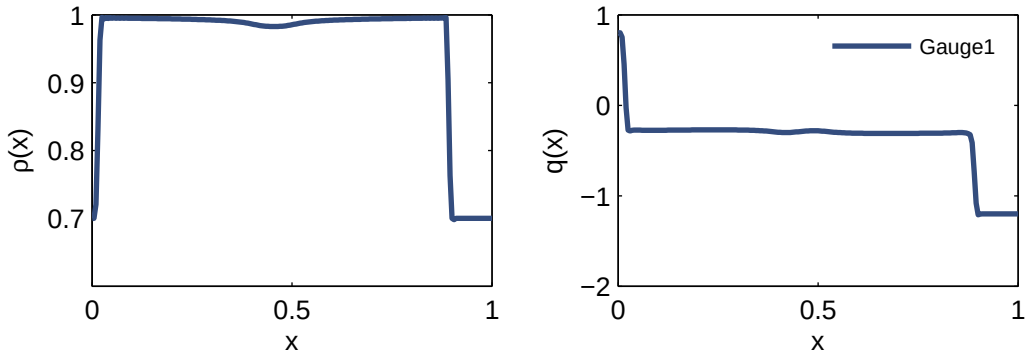
As above, we choose $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. It can be seen from Figure V.7 that the direct method performs well when the density is far away from congestion. However, the Gauge 1 method does not work well possibly due to the extra diffusion.

Remark. All the features described above are preserved when we apply the fully implicit method (V.28)(V.29)(implicit in both ρ and $\frac{q \otimes q}{\rho}$) to the direct and Gauge 1 methods. There is a little improvement in the accuracy but no major one.

Finally, we can compare the Gauge 1 and Gauge 2 methods. With the stencil of (V.53) in the same setting as that of example 2, the Gauge 2 method yields almost the same numerical result as the direct method (see Fig. V.8).

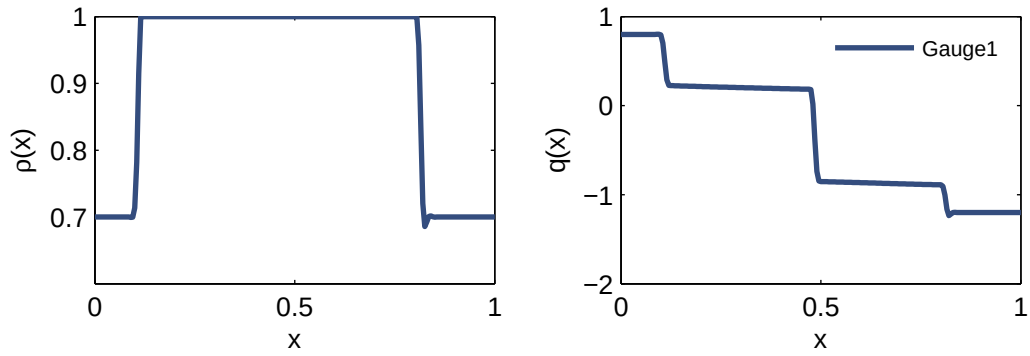


(a) $t = 0.0825$

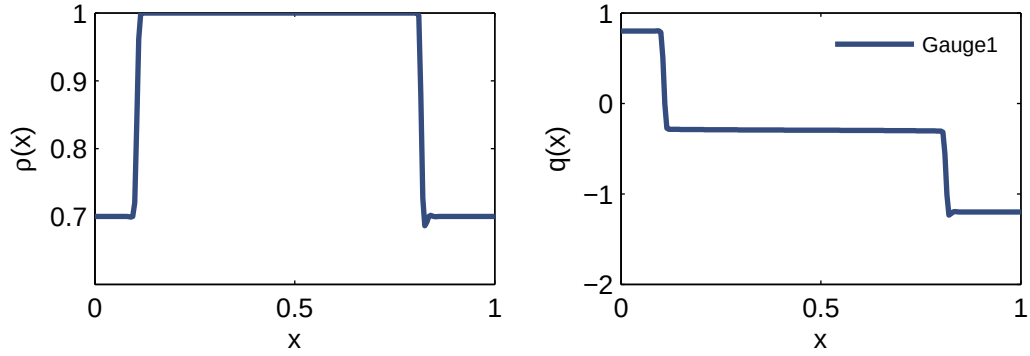


(b) $t = 0.1030$

Figure V.5: The Gauge 1 method for problem (P3) with $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}$ and $\Delta t = 5 \times 10^{-4}$. The left panels are for ρ, the right ones for q, both as a function of x.



(a) $t = 0.0810$



(b) $t = 0.0815$

Figure V.6: The Gauge 1 method for problem (P3) with $\varepsilon = 10^{-8}, \Delta x = 5 \times 10^{-3}$ and $\Delta t = 5 \times 10^{-4}$. The left panels are for ρ, the right ones for q, both as a function of x.

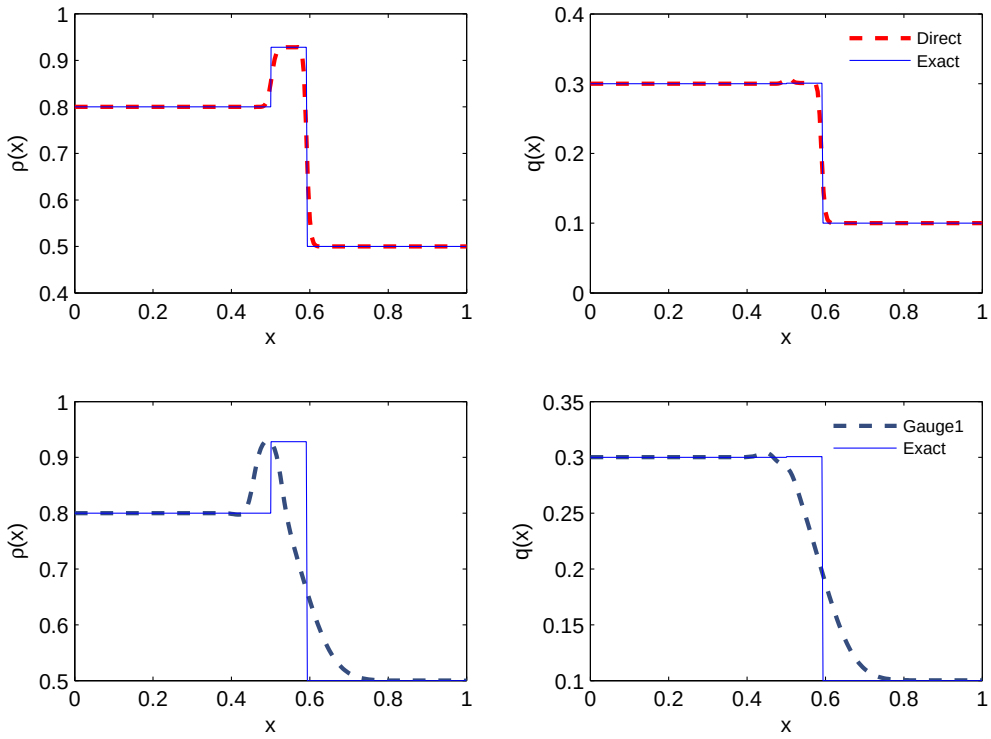


Figure V.7: The direct and the Gauge 1 method for problem (P4) at $t = 0.05$ with $\varepsilon = 10^{-4}, \Delta x = 5 \times 10^{-3}, \Delta t = 5 \times 10^{-4}$. The solid lines are the exact solution. The dashed curves are the numerical solutions. The left panels are for ρ, the right ones for q, both as a function of x.

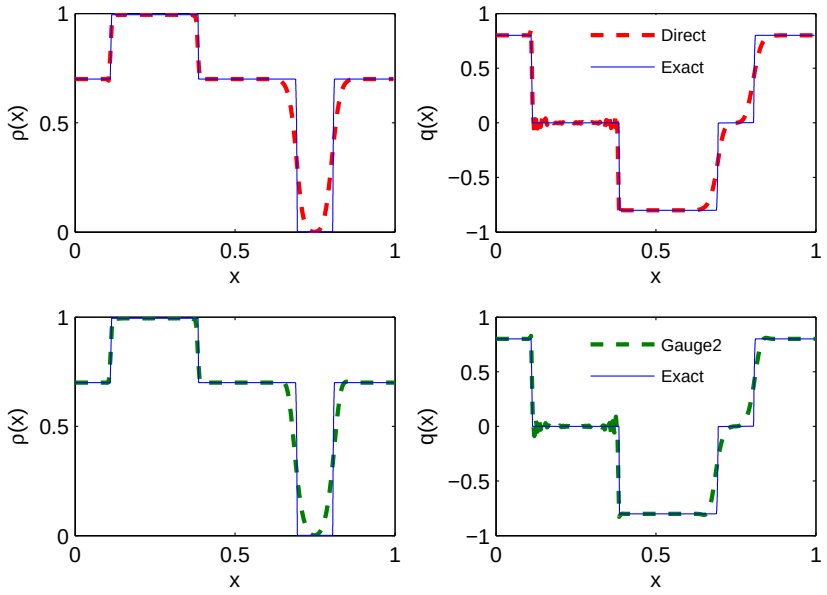


Figure V.8: The direct scheme and Gauge 2 methods with discretization (V.50) for problem (P2) at $t = 0.05$. The solid lines are the exact solution. The dashed curves are the numerical solutions. The left panels are for ρ, the right ones for q, both as a function of x.

6 Conclusion

In this chapter, we have studied the Euler system with a maximal density constraint. A small parameter ε was introduced to measure the stiffness of the constraint. As $\varepsilon \to 0$, the model gives rise to a two-phase model with congested regions (with maximal density) and uncongested regions (with density below the maximal density). One-dimensional solutions of this asymptotics problem have been carried out to provide informations on the interface conditions. However, it is not enough to characterize the whole dynamics. Therefore, we have devised asymptotic preserving numerical schemes, which are valid for all range of ε and thus are aimed to capture the asymptotic dynamics. Two numerical schemes have been considered and compared on one-dimensional test-cases. They both give good results. However, the first method shows some oscillations near the interface between congested and uncongested regions, while the second has much less oscillations but is more diffusive. Understanding their mathematical differences would enable to devise a method, which combine their advantages without their drawbacks.

A careful error analysis with respect to time and space steps is planned for the future. Two-dimensional simulations are also in current investigation. Finally, simulations of the model studied in chapter IV, with a supplementary geometric contraint on the speed of the flow, will be tackled.

A The one-dimensional Riemann problem

A.1 Rarefaction and shock waves

The **rarefaction waves** are continuous self similar solutions: $(\rho(x/t), q(x/t))$. Given a state $(\hat{\rho}, \hat{q})$, the states which can be connected to $(\hat{\rho}, \hat{q})$ by a rarefaction wave are those located on the integral curves $i_{\pm}^{\varepsilon}$ issued from $(\hat{\rho}, \hat{q})$. They are given by:

$$\rho'(s) = 1, \quad {i_{\pm}^{\varepsilon}}'(s) = \hat{u} \pm \sqrt{\varepsilon p'(\rho)}, \quad \rho(0) = \hat{\rho}, \quad i_{\pm}^{\varepsilon}(0) = \hat{q},$$

which is equivalent to ${i_{\pm}^{\varepsilon}}'(\rho) = \hat{u} \pm \sqrt{\varepsilon p'(\rho)}$, $i_{\pm}^{\varepsilon}(\hat{\rho}) = \hat{q}$ and then to equation (V.14). The graph of i_{-}^{ε} (resp. i_{+}^{ε}) is called the 1-integral curve (resp. the 2-integral curve). The following proposition provides several features of the integral curves.

Proposition 34. ***(Integral and rarefaction curves)***

1. *The 1-integral curve i_{-}^{ε} (resp. the 2-integral curve i_{+}^{ε}) is concave (resp. convex) as function of ρ and $i_{-}^{\varepsilon}(0) = i_{+}^{\varepsilon}(0) = 0$.*
2. *The limit of the integral curves as ε goes to zero is the union of the straight lines $\{q = \rho\hat{u}\}$ and $\{\rho = \rho^*\}$.*
3. *Suppose that the state $\hat{\rho}^{\varepsilon}$ is such that $\hat{\rho}^{\varepsilon} \to \rho^*$ and $\varepsilon p(\hat{\rho}^{\varepsilon}) \to \bar{p}$. For all $\rho < \hat{\rho}^{\varepsilon}$, we have:*
$$\left| \frac{i_{\pm}^{\varepsilon}(\rho)}{\rho} - \hat{u} \right| \leqslant \sqrt{\varepsilon} \int_0^{\hat{\rho}^{\varepsilon}} \frac{\sqrt{p'(u)}}{u} du = \underset{\hat{\rho}^{\varepsilon} \to \rho^*}{O}(\varepsilon^{\frac{1}{2\gamma}}).$$
4. *If $(\hat{\rho}, \hat{q})$ is a left state, the right states which can be connected to it by a rarefaction wave are those located on the 1-rarefaction curve $\left\{(\rho, i_{-}^{\varepsilon}(\rho)),\ \rho < \hat{\rho}\right\}$ or the 2-rarefaction curve $\left\{(\rho, i_{+}^{\varepsilon}(\rho)),\ \rho > \hat{\rho}\right\}$.*

The last point of this proposition stems from the compatibility conditions of the caracteristic speeds. The proof of this proposition is classical and omitted.

A **shock wave** is a discontinuity between two constant states, $(\hat{\rho}, \hat{q})$ and (ρ, q), travelling at constant speed σ. The states (ρ, q), which can be connected to $(\hat{\rho}, \hat{q})$ by a shock wave, are determined by the Rankine-Hugoniot conditions:

$$[q] = \sigma[\rho], \tag{V.58}$$

$$\left[\frac{q^2}{\rho} + \varepsilon p(\rho)\right] = \sigma[q], \tag{V.59}$$

where $[f] := f - \hat{f}$ for all quantities f. Easy computations show that system (V.58)-(V.59) is equivalent to system (V.15)-(V.16), which defines the Hugoniot functions $h_{\pm}^{\varepsilon}$. The graph of h_{-}^{ε} (resp. h_{+}^{ε}) is called the 1-Hugoniot curve (resp. the 2-Hugoniot curve). The following proposition provides several properties of the Hugoniot curves:

Proposition 35. ***(Hugoniot and shock curves)***

1. *The 1-Hugoniot function h_{-}^{ε} (resp. the 2-Hugoniot function h_{+}^{ε}) is concave (resp. convex) and $h_{-}^{\varepsilon}(0) = h_{+}^{\varepsilon}(0) = 0$.*

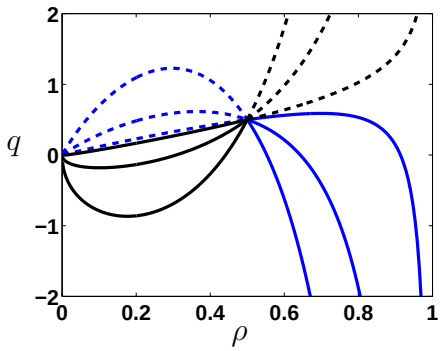


Figure V.9: Curves issued from $(\hat{\rho}, \hat{q}) = (0.5, 0.5)$ and $\varepsilon \in \{10, 1, 10^{-2}\}$. In dotted line : the rarefaction curves. In solid line : the shock curves. In blue : the 1-curves. In black : the 2-curves. As $\varepsilon \to 0$, the curves are tending to the straight lines $\{q = \hat{u}\rho\}$ and $\{\rho = \rho^*\}$. Parameters : $k = 2, \rho^* = 1$.

2. *The limits of the graphs of h_-^ε and h_+^ε when ε goes to zero are the union of the straight lines $\{q = \rho\hat{u}\}$ and $\{\rho = \rho^*\}$. This is true also when $\hat{\rho} = \hat{\rho}^\varepsilon$ depends on ε and tends to ρ^*.*
3. *If $(\hat{\rho}, \hat{q})$ is a left state, the right states which can be connected to it by an entropic shock wave are those located on the 1-shock curve $\left\{(\rho, h_-^\varepsilon(\rho)),\ \rho > \hat{\rho}\right\}$ or the 2-shock curve $\left\{(\rho, h_+^\varepsilon(\rho)),\ \rho < \hat{\rho}\right\}$.*

The first two points are deduced from straightforward computations. The last point of this proposition stems from entropic conditions (see [18]). Indeed, the Lax entropic conditions imply that a state (ρ, q) on the Hugoniot curve can be connected to $(\hat{\rho}, \hat{q})$ by a shock wave iff these inequalities hold:

$$\begin{aligned}\lambda_-(\hat{\rho}, \hat{q}) = \hat{u} - \sqrt{\varepsilon p'(\hat{\rho})} < s &= \hat{u} - \sqrt{\frac{\rho}{\hat{\rho}}}\sqrt{\frac{[\varepsilon p(\rho)]}{[\rho]}} \\ &= u - \sqrt{\frac{\hat{\rho}}{\rho}}\sqrt{\frac{[\varepsilon p(\rho)]}{[\rho]}} < \lambda_-(\rho, q) = u - \sqrt{\varepsilon p'(\rho)}.\end{aligned}$$

This reduces to the two following inequalities:

$$\hat{\rho}p'(\hat{\rho}) > \rho\frac{[p(\rho)]}{[\rho]}, \quad \rho p'(\rho) < \hat{\rho}\frac{[p(\rho)]}{[\rho]},$$

and the convexity of p yields the result.

In Fig. V.9, the behaviour of the rarefaction and shock curves is depicted.

A.2 Proofs of propositions 30, 31 and 32 (limit of solutions of the Riemann problem)

In all the following proofs, (ρ_ℓ, q_ℓ), $(\tilde{\rho}, \tilde{q})$, (ρ_r, q_r) will respectively denote the left, the intermediate and the right states involved in each different Riemann problems. The

solutions of the Riemann problem will consist of two waves : one wave connecting (ρ_ℓ, q_ℓ) to $(\tilde{\rho}, \tilde{q})$ and another wave connecting $(\tilde{\rho}, \tilde{q})$ to (ρ_r, q_r). The nature of the waves can be either a rarefaction or a shock wave. In the following, i_-^ε and h_-^ε (resp. i_+^ε and h_+^ε) will refer to the 1-curves (resp. 2-curves) issued from the left state (resp. right state). λ_-^ε (resp. λ_+^ε) will implicitly refer to the first (resp. second) characteristic speed of the left state (resp. right state). The characteristic speeds related to the intermediate state are denoted: $\tilde{\lambda}_-^\varepsilon$, $\tilde{\lambda}_+^\varepsilon$. The notations $[f]_\ell = \tilde{f} - f_\ell$ (resp. $[f]_r = \tilde{f} - f_r$) will denote the difference between the intermediate and the left (resp. the right) values of any quantity f. The following proposition provides how to obtain the intermediate states.

Proposition 36. *Let (ρ_ℓ, q_ℓ), (ρ_r, q_r), left and right states. Then the solutions of the Riemann problem related to (V.12)-(V.13) have the following forms:*

1. *If $u_\ell < u_r$, then for ε small enough, the intermediate state is the intersection point of the 1-integral curve issued from $(\rho_\ell, q_\ell, \bar{p}_\ell)$ and the 2-integral curve issued from $(\rho_r, q_r, \bar{p}_r)$. Besides, the intermediate density $\tilde{\rho}$ is lower than ρ_r and ρ_ℓ. This is valid even if ρ_ℓ or/and ρ_r tends to ρ^*.*
2. *If $u_\ell > u_r$, then for ε small enough, there are two subcases:*
   - *if $h_-^\varepsilon(\rho_r^\varepsilon) > q_r$ and $h_+^\varepsilon(\rho_\ell^\varepsilon) < q_\ell$, the intermediate state is the intersection point of the 1-Hugoniot curve issued from (ρ_ℓ, q_ℓ) and the 2-Hugoniot curve issued from (ρ_r, q_r).*
   - *if $h_-^\varepsilon(\rho_r^\varepsilon) < q_r$ (resp. $h_+^\varepsilon(\rho_\ell^\varepsilon) > q_\ell$), the intermediate state is the intersection point of the 1-Hugoniot curve (resp. 1-integral curve) issued from (ρ_ℓ, q_ℓ) and the 2-integral curve (resp. 2-Hugoniot curve) issued from (ρ_r, q_r).*
3. *If $u_r = u_\ell$ and $\rho_\ell < \rho_r$ (resp. $\rho_\ell > \rho_r$), then the intermediate state is the intersection point of the 1-Hugoniot curve (resp. 1-integral curve) issued from $(\rho_\ell, q_\ell, \bar{p}_\ell)$ and the 2-integral curve (resp. 2-Hugoniot curve) issued from $(\rho_r, q_r, \bar{p}_r)$. Besides, the intermediate density $\tilde{\rho}$ is the interval $[\rho_\ell, \rho_r]$ (resp. $[\rho_r, \rho_\ell]$). This is valid even if ρ_ℓ or/and ρ_r tends to ρ^*.*

This proposition results from propositions 34 and 35. The proof is omitted here. Note that the nature of the curves (integral or Hugoniot curve) implies the nature of the waves involved in the Riemann problem.

A.2.1 Proof of proposition 30

Proof. 1. From proposition 36, the solution for small ε consists of two rarefaction waves. The intermediate density solves equation $i_-^\varepsilon(\tilde{\rho}) = i_+^\varepsilon(\tilde{\rho})$, that is:

$$\tilde{\rho}u_r \pm \tilde{\rho}\sqrt{\varepsilon}[P(\rho)]_r = \tilde{\rho}u_\ell \pm \tilde{\rho}\sqrt{\varepsilon}[P(\rho)]_\ell.$$

Since $\tilde{\rho}$ is lower than ρ_ℓ and ρ_r (and then $[P(\rho)]_r$ and $[P(\rho)]_\ell$ are bounded) and $u_r - u_\ell$ is not zero, it is easy to deduce that the limit solution of this equation is $\tilde{\rho} = 0$, which defines a vacuum state. Besides, λ_-^ε and $\tilde{\lambda}_-^\varepsilon$ tend to u_ℓ (resp. $\tilde{\lambda}_+^\varepsilon$ and λ_+^ε tends to u_r). Therefore, the limit waves are contact waves.

2. From proposition 36, the solution for small ε consists of two shock waves (since $\lim h_+^\varepsilon(\rho_\ell) = \rho_\ell u_r$ and $\lim h_-^\varepsilon(\rho_\ell) = \rho_r q_\ell$). We have $h_{-,\ell}^\varepsilon(\tilde{\rho}) = h_{+,r}^\varepsilon(\tilde{\rho})$, that is:

$$\tilde{\rho}u_\ell - \sqrt{\frac{\tilde{\rho}}{\rho_\ell}}\sqrt{[\rho]_\ell[\varepsilon p(\rho)]_\ell} = \tilde{\rho}u_r + \sqrt{\frac{\tilde{\rho}}{\rho_r}}\sqrt{[\rho]_r[\varepsilon p(\rho)]_r}.$$

which yields

$$(u_\ell - u_r) = \sqrt{\frac{[\rho]_\ell[\varepsilon p(\rho)]_\ell}{\rho_\ell \tilde{\rho}}} + \sqrt{\frac{[\rho]_r[\varepsilon p(\rho)]_r}{\rho_r \tilde{\rho}}},$$

Since $u_\ell - u_r$ is different from zero, the limit intermediate pressure: $\lim[\varepsilon p(\rho)]_\ell = \lim[\varepsilon p(\rho)]_r = \lim \varepsilon p(\tilde{\rho}^\varepsilon)$ is not zero. Thus $\tilde{\rho} \to \rho^*$ as $\varepsilon \to 0$. Finally, the limit values of the two shock speeds can be easily infered from equation (V.16).

3. Supposing that $\rho_\ell < \rho_r$, then the intermediate density satisfies $\rho_\ell \leqslant \tilde{\rho}^\varepsilon \leqslant \rho_r$ and consequently $\lambda_-^\varepsilon, \tilde{\lambda}_\pm^\varepsilon, \lambda_+^\varepsilon \to u_\ell = u_r$ as ε goes to zero, which yields a unique contact wave. ∎

A.2.2 Proof of proposition 31

Proof. 1. From proposition 36, the solution for small ε consists of two rarefaction waves and it can be easily checked that the limit intermediate density is zero (thanks to proposition 34, point 3, $\sqrt{\varepsilon}[P(\rho)]_\ell = O(\varepsilon^{\frac{1}{2\gamma}})$). Since ρ_ℓ tends to ρ^*, we have $\lim \lambda_-^\varepsilon = -\infty$ and $\lim \tilde{\lambda}_-^\varepsilon = u_\ell$. The limit of the 2-rarefaction wave is a contact wave (since $\lim \tilde{\lambda}_+^\varepsilon = \lim \lambda_+^\varepsilon = u_r$). Let us look at the limit of the 1-rarefaction wave.

For each possible speed $s \in [\lambda_-^\varepsilon, u_\ell]$ of the rarefaction wave connecting the left state to the intermediate state, we have:

$$s = \lambda_-(\rho(s), q(s)) = \frac{q(s)}{\rho(s)} - \sqrt{\varepsilon p'(\rho(s))}. \tag{V.60}$$

The state $(\rho(s), q(s))$ belongs to the integral curve and then for $\rho(s) < \rho_\ell$ and according to proposition 34, point 3, $q(s)/\rho(s)$ tends to u_ℓ. If $s \neq u_\ell$, equation (V.60) implies that $\rho(s)$ has to tend to ρ^* and that $\lim \varepsilon p(\rho(s)) = 0$ since $\lim \varepsilon p'(\rho(s))$ is finite. This yields the definition of a declustering wave, given in definition 2.

2. We have $\lim h_-^\varepsilon(\rho_r) = \rho_r u_\ell > q_r$. According to proposition 36, in order to have a solution which is the limit of two shock waves, we have to prescribe:

$$h_+^\varepsilon(\rho_\ell) = \rho_\ell u_r + \sqrt{\varepsilon}\sqrt{\frac{\rho_\ell}{\rho_r}}\sqrt{[\rho]_r^\ell [p(\rho)]_r^\ell} < q_\ell = \rho_\ell u_\ell,$$

and in the limit:

$$u_r + \sqrt{\frac{(\rho_* - \rho_r)}{\rho_r \rho_*}}\sqrt{\bar{p}_\ell} \leqslant u_\ell.$$

If $h_+^\varepsilon(\rho_\ell) > q_\ell$, the solution is the limit of a 1-rarefaction wave and a 2-shock wave. According to this discussion, we have to consider two cases but we will see that the limits of the two cases are the same.

The first case corresponds to the limit of two shock waves. The intermediate density is then greater than the left one and so tends to ρ^* too. Besides we have:

$$\tilde{\rho} u_\ell - \sqrt{\frac{\tilde{\rho}}{\rho_\ell}}\sqrt{[\rho]_\ell[\varepsilon p(\rho)]_\ell} = \tilde{\rho} u_r + \sqrt{\frac{\tilde{\rho}}{\rho_r}}\sqrt{[\rho]_r[\varepsilon p(\rho)]_r}, \tag{V.61}$$

and then by dividing by $\tilde{\rho}$, we get:

$$u_\ell - \sqrt{\frac{[\rho]_\ell[\varepsilon p(\rho)]_\ell}{\tilde{\rho}\rho_\ell}} = u_r + \sqrt{\frac{[\rho]_r[\varepsilon p(\rho)]_r}{\tilde{\rho}\rho_r}} < u_\ell.$$

The last inequality implies that the limit of $\varepsilon p(\tilde{\rho})$ is finite. We denote it by $\bar{p}$. Taking the limit in the equality (V.61), we obtain

$$u_\ell = u_r + \sqrt{\frac{(\rho^* - \rho_r)}{\rho^* \rho_r}}\sqrt{\bar{p}}.$$

Hence we find the value of $\bar{p}$ as in proposition 31. The propagation speeds are easily deduced from the limit $\varepsilon \to 0$ in (V.16).

If the solution is the limit of a combination of a rarefaction wave and a shock wave, then the intermediate state satisfies:

$$\tilde{\rho} u_\ell - \tilde{\rho}\sqrt{\varepsilon}[P(\rho)]_\ell = \tilde{\rho} u_r + \sqrt{\frac{\tilde{\rho}}{\rho_r}}\sqrt{[\rho]_r [\varepsilon p(\rho)]_r}.$$

If the intermediate density $\tilde{\rho}^\varepsilon$ did tend to ρ^*, then we would obtain $u_\ell = u_r$ (thanks to proposition 34), which is impossible. Thus the intermediate density tends to ρ^* and the previous expression tends to:

$$\rho^* u_\ell = \rho^* u_r + \sqrt{\frac{\rho^*}{\rho_r}}\sqrt{(\rho^* - \rho_r)\bar{p}},$$

which yields the expected result. Since the pressure is positive, $\tilde{\lambda}^\varepsilon_-$ tends to $-\infty$ which implies that the rarefaction wave turns into a shock wave with infinite propagation speed, i.e. a declustering wave.

3. From proposition 36, the solution is the limit of a combination of a 1-rarefaction wave and a 2-shock wave. Using the fact that $\tilde{\rho} u_\ell = \tilde{\rho} u_r$, the intermediate state $(\tilde{\rho}, \tilde{q})$ satisfies:

$$\tilde{\rho}\varepsilon[P(\rho)]_\ell = \sqrt{\frac{\rho}{\rho_\ell}}\sqrt{[\rho]_r [\varepsilon p(\rho)]_r}.$$

So $[\rho]_r[\varepsilon p(\rho)]_r \to 0$ (since $\varepsilon P(\tilde{\rho}) < \varepsilon P(\rho_\ell)$ tends to zero) and either $\tilde{\rho}$ tends to ρ_r or $\varepsilon p(\tilde{\rho})$ tends to zero. Actually, whatever the limit value of the intermediate density, the intermediate state disappears. Indeed, let us consider all the possible cases.

Either $\tilde{\rho} \to \rho_r$. Then the 2-shock wave disappears and the 1-rarefaction wave tends to the sum of a declustering wave and a contact wave, which can be proven as in the case 1 of this proof.

Or $\lim \tilde{\rho} \in]\rho_r, \rho^*[$. It is easy to check that the 2-shock wave becomes a contact wave and the intermediate state disappears since $\tilde{\lambda}^\varepsilon_+, \tilde{\lambda}^\varepsilon_- \to u_\ell$. Like in the case 1 of this proof, the 1-rarefaction wave leads to declustering wave and a contact wave, which superimposes on the one coming from the 2-shock wave.

Or $\lim \tilde{\rho} = \rho^*$. Then we know that $\varepsilon p(\tilde{\rho})$ tends to zero. Thus the 1-rarefaction wave tends to a declustering wave and the 2-shock wave tends to a contact wave.

In all the cases, the limit solution is the one given in proposition 31. ∎

A.2.3 Proof of proposition 32

Proof. 1. The arguments are similar to those used in the proof of the first case of the previous proposition (appendix A.2.2).

2. From proposition 36, the solution is the limit of two shock waves if for small ε, we have the following inequalities:

$$h_+^\varepsilon(\rho_\ell^\varepsilon) = \rho_\ell^\varepsilon u_r + \sqrt{\varepsilon}\sqrt{\frac{\rho_\ell^\varepsilon}{\rho_r^\varepsilon}}\sqrt{[\rho]_r^\ell [p(\rho)]_r^\ell} < q_\ell = \rho_\ell^\varepsilon u_\ell,$$
$$h_-^\varepsilon(\rho_r^\varepsilon) = \rho_r^\varepsilon u_\ell + \sqrt{\varepsilon}\sqrt{\frac{\rho_r^\varepsilon}{\rho_\ell^\varepsilon}}\sqrt{[\rho]_r^\ell [p(\rho)]_r^\ell} > q_r = \rho_r^\varepsilon u_r.$$

Here, these inequalities are always satisfied : their limit is $u_r < u_\ell$ since $\varepsilon[p(\rho)]_r^\ell$ is bounded and $[\rho]_r^\ell \to 0$. The intermediate density $\tilde{\rho}$ satisfies:

$$u_\ell - \sqrt{\frac{[\rho]_\ell[\varepsilon p(\rho)]_\ell}{\tilde{\rho}\rho_\ell}} = u_r + \sqrt{\frac{[\rho]_r[\varepsilon p(\rho)]_r}{\tilde{\rho}\rho_r}}.$$

Therefore, $\varepsilon p(\tilde{\rho})$ cannot be bounded (which would imply $u_\ell = u_r$ since $[\rho]_\ell$ and $[\rho]_r$ tend to zero). So $\varepsilon p(\tilde{\rho})$ tends to $+\infty$ as $\varepsilon \to 0$. From the Rankine-Hugoniot relations, we have

$$[\rho q + \varepsilon p(\rho)]_\ell [\rho]_\ell = [q]_\ell^2,$$
$$[\rho q + \varepsilon p(\rho)]_r [\rho]_r = [q]_r^2.$$

Taking the limit of their quotient, we obtain:

$$\lim \frac{[\varepsilon p(\rho)]_\ell [\rho]_\ell}{[\varepsilon p(\rho)]_r [\rho]_r} = \left(\frac{[q]_\ell}{[q]_r}\right)^2. \tag{V.62}$$

Besides, we have:

$$\frac{[\varepsilon p(\rho)]_\ell [\rho]_\ell}{[\varepsilon p(\rho)]_r [\rho]_r} \underset{\varepsilon\to 0}{\sim} \frac{[\rho]_\ell}{[\rho]_r} \underset{\varepsilon\to 0}{\sim} \frac{\rho^* - \rho_\ell}{\rho^* - \rho_r}$$

where the last equivalence results from the fact that $(\rho^* - \tilde{\rho}) = o(\varepsilon^{\frac{1}{\gamma}})$ and $(\rho^* - \rho_{\ell,r}) = O(\varepsilon^{\frac{1}{\gamma}})$. Finally, we have:

$$\frac{\rho^* - \rho_\ell}{\rho^* - \rho_r} = \left(\frac{\varepsilon p(\rho_r)}{\varepsilon p(\rho_\ell)}\right)^{\frac{1}{\gamma}} \underset{\varepsilon\to 0}{\to} \left(\frac{\bar{p}_r}{\bar{p}_\ell}\right)^{\frac{1}{\gamma}}.$$

This last result, combined with eq. (V.62), provides an equation for the intermediate momentum.

3. Let us suppose that $\rho_\ell^\varepsilon > \rho_r^\varepsilon$. The intermediate density satisfies $\rho_r^\varepsilon < \tilde{\rho}^\varepsilon < \rho_\ell^\varepsilon$ and so tends to ρ^*. The intermediate momentum $\tilde{q}$ tends to $\rho^* u_\ell$. The 1-rarefaction wave tends to a shock wave with infinite propagation speed (since $\tilde{\lambda}_-^\varepsilon$ and λ_-^ε tend to $-\infty$), i.e. a declustering wave. We have now to determine $\bar{p}$. We have

$$\sqrt{\frac{\tilde{\rho}}{\rho_r}}\sqrt{[\rho]_r[\varepsilon p(\rho)]_r} = -\sqrt{\varepsilon}[P(\rho)]_\ell.$$

Since $\tilde{\rho} - \rho_\ell \to 0$, we have $-[P(\rho)]_\ell \sim (\rho_\ell - \tilde{\rho})P'(\rho_\ell)$. We have $P'(\rho_\ell) \sim C(\rho^* - \rho_\ell)^{-\frac{\gamma+1}{2}}$. Then we have $P'(\rho_\ell) \sim C(\varepsilon^{\frac{1}{\gamma}})^{-\frac{\gamma+1}{2}} = C\varepsilon^{-\frac{1}{2}-\frac{1}{2\gamma}}$ (since $(\rho_* - \rho_\ell) = O(\varepsilon^{\frac{1}{\gamma}})$). Besides $(\rho_\ell - \tilde{\rho}) \leqslant (\rho_* - \rho_r) = O(\varepsilon^{\frac{1}{\gamma}})$ and $[\rho]_r = O(\varepsilon^{\frac{1}{\gamma}})$ and , so we have

$$[\varepsilon p(\rho)]_r = \frac{\rho_r}{\tilde{\rho}}\frac{(-\sqrt{\varepsilon}[P(\rho)]_\ell)^2}{[\rho]_r} = O(\varepsilon^{\frac{1}{\gamma}}).$$

So $\varepsilon p(\tilde{\rho})$ tends to $\bar{p}_r$. The 2-shock wave disappears (since the intermediate and the right state are identical). Then the limit solution consists of an instantaneous propagation of the right state. ∎

Bibliography

[1] D.M. Anderson, G.B. McFadden, A.A. Wheeler, Diffuse-interface methods in fluid mechanics, *Annu. Rev. Fluid. Mech.*, 30(1) :139-165, 1998.

[2] N. Bellomo, C. Dogbe, On the modelling crowd dynamics from scaling to hyperbolic macroscopic models, *Math. Models Methods Appl. Sci.*, 18 :1317-1345, 2008.

[3] F. Berthelin, Existence and weak stability for a pressureless model with unilateral constraint, *Math. Models Methods Appl. Sci.*, 12(2) :249-272, 2002.

[4] F. Berthelin, P. Degond, M. Delitala, M. Rascle, A model for the formation and the evolution of traffic jams, *Arch. Rational Mech. Anal.*, 187 :185-220, 2008.

[5] F. Berthelin, P. Degond, V. Le Blanc, S. Moutari, M. Rascle, J. Royer, A traffic-flow model with constraints for the modeling of traffic jams, *Math. Models Methods Appl. Sci.*, 18 :1269-1298, 2008.

[6] F. Bouchut, On zero pressure gas dynamics, Advances in kinetic theory and computing: selected papers, *Ser. Adv. Math. Appl. Sci.*, 22 :171-190, 1994.

[7] F. Bouchut, Y. Brenier, J. Cortes, J.-F. Ripoll, A hierarchy of models for two-phase flows, *J. Nonlinear Sci.*, 10 :639-660, 2000.

[8] Y. Brenier, E. Grenier, Sticky particles and scalar conservation laws, *SIAM J. Num. Anal.*, 35(6) :2317-2328, 1998.

[9] P. Degond, M. Delitala, Modelling and simulation of vehicular traffic jam formation, *Kinetic Related Models*, 1(2) :279-293, 2009.

[10] P. Degond, S. Jin, J.-G. Liu, Mach-number uniform asymptotic-preserving gauge schemes for compressible flows, *Bulletin of the institute of Mathematics, Academia Sinica*, 2(4) :851-892 2007.

[11] P. Degond, L. Navoret, R. Bon, D. Sanchez, Congestion in a Macroscopic Model of Self-driven Particles Modeling Gregariousness, *J. Stat. Phys.*, 138(1-3) :85-125, 2010.

[12] P. Degond, M. Tang, Asymptotic preserving method for the incompressible low mach number limit of the Isentropic Euler equation, in prep., 2009.

[13] H. Guillard, C. Viozat, On the behaviour of upwind schemes in the low Mach number limit, *Computers and fluids*, 28(1) :63-86, 1999.

[14] F.H. Harlow, A.A. Amsden, A numerical fluid dynamics calculation method for all flow speeds, *J. Comput. Phys.*, 8(2) :197-213, 1971.

[15] H. Hattori, The Riemann problem for a van der Waals fluid with entropy rate admissibility criterion, Isothermal case, *Archive for Rational Mechanics and Analysis*, 92(3) :247-263, 1986.

[16] C.W. Hirt, B.D. Nichols, Volume of fluid (VOF) method for the dynamics of free boundaries, *J. Comp. Phys.*, 39(1) :201-225, 1981.

[17] R. Klein, Semi-implicit extension of a Godunov-type scheme based on low Mach number asymptotics I: one-dimensional flow, *J. Comput. Phys.*, 121(2) :213-237, 1995.

[18] R.J. LeVeque, *Numerical Methods for Conservation Laws*. Birkhäuser-Verlag, Basel, 1990.

[19] P.L. Lions, N. Masmoudi, On a free boundary barotropic model, *Annales de l'Institut Henri Poincare/Analyse non lineaire*, 16(3) :373-410, 1999.

[20] C.D.Munz, S. Roller, R. Klein, K.J. Geratz, The extension of incompressible flow solvers to the weakly compressible regime, *Computers and Fluids*, 32(2) :173-196, 2003.

[21] S. Osher, R.P. Fedkiw, *Level set methods and dynamic implicit surfaces*, Springer, 2003.

[22] S. V. Patankar, *Numerical heat transfer and fluid flow*, New York: McGraw-Hill, 1980.

[23] G. Tryggvason, B. Bunner, A. Esmaeeli, D. Juric, N. Al-Rawahi, W. Tauber, J. Han, S. Nas, Y.-H. Jan, A front-tracking method for the computations of multiphase flow, *J. Comp. Phys.*, 169(2) :708-759, 2001.

Conclusion et perspectives

Dans cette thèse, nous avons abordé plusieurs problèmes asymptotiques dans leurs aspects numériques ou analytiques.

Au chapitre I, nous avons proposé deux méthodes particulaires (PIC) préservant la limite quasi-neutre pour le système Vlasov-Poisson. Les deux méthodes sont basées sur une résolution semi-implicite des trajectoires des particules et sur une reformulation de l'équation de Poisson ne dégénérant pas lorsque la longueur de Debye devient très faible. Plusieurs cas tests ont montré que contrairement à la méthode explicite, ces méthodes sont stables lorsque les échelles de quasi-neutralité ne sont pas résolues, c'est-à-dire lorsque le pas de temps est plus grand que la période plasma et le pas d'espace plus grand que la longueur de Debye. De plus, la consistance asymptotique des deux schémas a été confirmée par un cas test raide d'expansion de plasma. La stabilité et la consistance asymptotique de ces schémas permettent donc l'utilisation de grands pas de temps et d'espace et donc de réduire les coûts des simulations numériques.

Bien que les différents cas tests numériques fournissent une solide validation de la consistance asymptotique des schémas, une preuve mathématique pourra être entreprise. Par ailleurs, les méthodes se sont révélées fortement dissipatives notamment lorsque les échelles de quasi-neutralité sont résolues. Ceci est certainement dû au bruit numérique lors de l'assignation des quantités macroscopiques à partir des particules. Des méthodes de réduction de bruit pourront donc être envisagées. Pour améliorer la précision de ces méthodes, nous pourrons aussi mettre au point des schémas avec une précision d'ordre deux en temps pour intégrer les trajectoires des particules. Enfin, lorsque les électrons ont une masse très faible, les vitesses atteintes par les électrons peuvent devenir très grandes et la condition CFL devenir très contraignante et il serait donc bon de mettre au point des méthodes numériques permettant de lever cette contrainte.

Aux chapitres II et III, nous nous sommes intéressés à la description macroscopique de systèmes de particules avec des interactions d'alignement de type Vicsek et une contrainte géométrique sur le module des vitesses des particules.

Dans ce contexte, nous avons tout d'abord étudié un modèle de Vicsek à deux populations. Un modèle a été dérivé lorsque les fréquences d'alignement au sein d'une même population sont plus élevées que les fréquences d'échanges entre les deux populations. Cette étude n'est pas terminée : la stabilité des équilibres ainsi que les propriétés mathématiques du modèle obtenu (hyperbolicité) restent encore à explorer. De plus, lorsque les fréquences d'échanges sont plus importantes que les fréquences d'interactions d'alignement au sein des populations, l'existence même d'équilibres est encore à démontrer dans le cas général. Nous pourrons aussi étudier les modèles à deux populations avec des variantes du modèle de Vicsek (ajout d'un angle de vision, fréquence d'interaction dépendant de la densité locale,...).

Dans le chapitre III, différents schémas numériques pour le modèle macroscopique de Vicsek ont été comparés à la dynamique particulaire. Le schéma basé sur la relaxation de la contrainte géométrique sur la vitesse concorde avec les simulations particulaires. Ceci fournit une validation du modèle macroscopique de Vicsek lorsque la densité des particules est élevée.

Ce travail offre plusieurs perspectives. Tout d'abord, nous pourrons chercher à justifier mathématiquement la concordance entre le schéma de relaxation et les simulations particulaires. Les simulations bidimensionnelles sont en cours d'étude par Sébastien Motsch et des difficultés surviennent notamment lorsque des domaines de faibles densités apparaissent. La gestion numérique et théorique de ces faibles densités pourra être explorée. En appendice, une solution en vortex pour le modèle macroscopique a été exhibée mais il semblerait qu'elle soit numériquement instable. La recherche de conditions aux limites adéquates pour la stabiliser est en cours. Enfin, ce travail permettra de confronter le système obtenu au chapitre II à la dynamique particulaire à deux vitesses.

Aux chapitres IV et V, nous nous sommes intéressés à la prise en compte d'une contrainte de densité maximale.

Le chapitre IV a été consacré à l'étude d'un modèle macroscopique décrivant des effets de congestion dans un système de particules. Il a été dérivé d'une dynamique particulaire avec une force répulsive à courte portée devenant prédominante lorsque la densité locale devient proche de la densité maximale. Le modèle macroscopique asymptotique comporte deux phases : une phase incompressible dans laquelle la densité maximale est atteinte et une phase compressible pour les densités plus faibles. La dynamique de l'interface entre les deux phases a été en partie obtenue par l'analyse d'un problème de Riemann.

Nous avons fourni une solution unidimensionnelle de la collision de deux domaines de densité maximale, néanmoins la géométrie de la collision bidimensionnelle n'a pas pu être obtenue. Des recherches dans cette voie pourront être engagées. Au niveau du modèle microscopique, des simulations numériques ont été présentées et montrent la formation d'agrégats. L'étude de paramètre va être poursuivie afin de déterminer quelles sont les conditions pour obtenir de tels phénomènes collectifs. De plus, l'alignement n'est pas observé dans ces agrégats contrairement à ce qui est obtenu au niveau macroscopique. La fermeture monocinétique du modèle n'est donc peut-être pas appropriée pour décrire ce modèle particulaire. D'autres asymptotiques pourront donc être envisagées. Enfin, nous pouvons penser au raffinement éventuel du modèle microscopique : par exemple l'ajout d'un angle de vision sur les interactions, l'ajout de force d'alignement de type Vicsek,...

Enfin, dans le chapitre V, nous avons proposé des schémas numériques permettant de capturer les transitions entre les phases compressibles et les phases incompressibles obtenues comme limite asymptotique d'un système d'Euler avec une contrainte de densité maximale portée par une pression singulière. Ces schémas présentent l'intérêt de générer les deux dynamiques par l'intermédiaire d'un seul schéma. De plus, la contrainte de densité maximale est satisfaite numériquement. Bien que donnant des résultats consistants avec la limite de congestion, les deux schémas proposés ont quelques imperfections. Un des deux schémas présente des oscillations sur la vitesse dans les domaines incompressibles de densité maximale et le second est très diffusif dans les domaines compressibles.

La preuve du caractère absolument consistant des schémas est encore incomplète du fait de la difficulté à traiter les interfaces entre les deux domaines. Par ailleurs, des anal-

yses d'erreurs seront effectuées et les extensions bidimensionnelles sont en cours. Pour affiner ces méthodes, nous pourrons aussi mettre au point des schémas d'ordre deux en temps et en espace. Finalement, cette étude combinée avec les résultats obtenus pour le modèle de Vicsek pourra permettre la simulation numérique du modèle étudié au chapitre IV et ainsi déterminer numériquement la géométrie bidimensionnelle des collisions entre deux régions de densité maximale.

Dans cette thèse, nous avons notamment étudié plusieurs modèles pour le déplacement collectif d'animaux. Ces modèles n'ont pas a priori la validité des modèles utilisés en physique des plasmas et il nous semble donc important de discuter leur pertinence biologique et leur utilité en tant qu'outils mathématiques. Les différentes études asymptotiques menées dans cette thèse apportent quelques réponses mais sans aucun doute, pour déterminer les domaines de validité des différents modèles, ils pourront (et devront) être confrontrés à des mesures expérimentales.

Title: Asympotical and Numerical Methods for Plasma Physics Problems and Social Interactions Models

---

Abstract:

In this thesis, we devise analytical and numerical methods for capturing the asymptotic dynamics of plasma physics problems and collective movement models for animal populations. In the first part, we present a Particle-In-Cell numerical method for the Vlasov-Poisson system that is asymptotic preserving for the quasineutral limit. In the second part, we study the macroscopic limit of a Vicsek model that describes alignement interactions among two populations: a moving population and a steady one. Then we select a numerical scheme for capturing the solutions of the macroscopic Vicsek model corresponding to the underlying particle dynamics. The third part is dedicated to the incompressible-compressible transitions that appear in a macroscopic model for collective displacements with congestion effects. Asymptotic preserving numerical schemes for the congestion limit are then built for the Euler system with a maximal density constraint.

Auteur : Laurent Navoret

Titre : Méthodes asymptotico-numériques pour des problèmes issus de la physique des plasmas et de la modélisation des interactions sociales

Directeurs de thèse : Pierre Degond & David Sanchez

Date et lieu de soutenance : le 30 juin 2010 à l'Université Toulouse 3 - Paul Sabatier

---

Résumé :

Dans cette thèse, nous développons des méthodes analytiques et numériques pour capturer les dynamiques asymptotiques de problèmes issus de la physique des plasmas et de la modélisation des mouvements collectifs dans les populations animales. Dans une première partie, nous présentons une méthode numérique Particle-In-Cell (PIC) pour le système Vlasov-Poisson préservant l'asymptotique quasi-neutre. Dans une seconde partie, nous étudions la limite macroscopique d'un modèle de Vicsek décrivant des interactions d'alignement entre deux populations, une population à l'arrêt et une population en mouvement. Nous sélectionnons ensuite un schéma numérique pour capturer les solutions du modèle macroscopique de Vicsek correspondant à la dynamique particulaire sous-jacente. La troisième partie est dédiée à l'étude des transitions compressible-incompressible apparaissant sous l'effet d'une contrainte de congestion dans un modèle macroscopique de déplacement collectif. Des schémas numériques préservant l'asymptotique de congestion sont ensuite mis au point pour le système d'Euler avec une contrainte de densité maximale.

Mots Clés : Limite quasi-neutre, Schéma numérique préservant l'asymptotique, Déplacement collectif d'animaux, Modèle d'alignement, Limite hydrodynamique, Congestion, Problème de Riemann, Méthode volume fini

---

Discipline administrative : Mathématiques

Intitulé et adresse du laboratoire : Institut de Mathématiques de Toulouse, Université Toulouse 3 - Paul Sabatier, 31062 Toulouse, France